

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:23:54 ; Search time 738.572 Seconds
(without alignments)
16735.564 Million cell updates/sec

Title: US-09-974-546C-83

Perfect score: 2088

Sequence: 1 gacctaataatagaggt.....tactacaacacctctaacgc 2088

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004as:*
13: Geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2087.6	100.0	2088	3	Aaz87583
2	2076	99.4	2087	4	Aas04000
3	1961.6	93.9	2506	3	Aaz87584
4	1950	93.4	2505	4	Aas04001
5	275.4	13.2	757	2	Aav16883
6	275.4	13.2	757	2	Aax26018
7	275.4	13.2	757	2	Aax26018
8	275.4	13.2	757	2	Aax26018
9	115.6	5.5	573	3	Aas03722
10	115.6	5.5	4051	5	Aa16415
11	115.2	5.5	430	6	Abt07086
12	115.2	5.5	430	8	Abx72964
13	115.2	5.5	1167	5	Aas71702
14	115.2	5.5	1167	5	Aas92271
15	115.2	5.5	1167	5	Aas75366
16	115.2	5.5	1167	5	Aas79702
17	115.2	5.5	1167	5	Aas84973
18	115.2	5.5	1205	5	Aas81380
19	115.2	5.5	1205	5	Aas72059
20	115.2	5.5	1205	5	Aas71116

21	115.2	5.5	1364	5	Aas71114	DNA	encod
22	115.2	5.5	1364	5	Aas72053	DNA	encod
23	115.2	5.5	1548	5	Aas71240	DNA	encod
24	115.2	5.5	1719	5	Aas70932	DNA	encod
25	115.2	5.5	1719	5	Aas91719	DNA	encod
26	115.2	5.5	1719	5	Aas69594	DNA	encod
27	115.2	5.5	1719	5	Aas73101	DNA	encod
28	115.2	5.5	1719	5	Aas83831	DNA	encod
29	115.2	5.5	2160	5	Aas64322	DNA	encod
30	115.2	5.5	2341	5	Aas85521	DNA	encod
31	115.2	5.5	2385	5	Aas81821	DNA	encod
32	115.2	5.5	2457	5	Aas73106	DNA	encod
33	115.2	5.5	2523	5	Aas74903	DNA	encod
34	115.2	5.5	2580	5	Aas78969	DNA	encod
35	115.2	5.5	2648	5	Aas71603	DNA	encod
36	115.2	5.5	2685	5	Aas73165	DNA	encod
37	115.2	5.5	2686	5	Aas77246	DNA	encod
38	115.2	5.5	2748	5	Aas71650	DNA	encod
39	115.2	5.5	2748	5	Aas76929	DNA	encod
40	115.2	5.5	2748	5	Aas72643	DNA	encod
41	115.2	5.5	2748	5	Aas76709	DNA	encod
42	115.2	5.5	2748	5	Aas66493	DNA	encod
43	115.2	5.5	2748	5	Aas78788	DNA	encod
44	115.2	5.5	2955	5	Aas71125	DNA	encod
45	115.2	5.5	3048	5	Aas92377	DNA	encod

ALIGNMENTS

RESULT 1

Aaz87583

ID Aaz87583 standard; DNA; 2088 BP.

XX Aaz87583;

AC Aaz87583;

DT 19-APR-2000 (first entry)

XX

DE Prostate disease marker UC Band #28.

XX

Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer; benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection; diagnosis; PCR primer; ss.

OS Homo sapiens.

XX

PN WO9964631-A1.

XX

PD 16-DEC-1999.

XX

PF 11-JUN-1999; 99WO-US013151.

XX

PR 12-JUN-1999; 98US-00097199.

XX

PA (UROC-) UROCOR INC.

XX

PI An G, O'hara SM, Ralph D, Veltri RW;

XX

DR WPI; 2000-116557/10.

XX

DR P-PSDB; AAY59295.

XX

PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate, breast and bladder cancer.

XX

PS Claim 1; Page 182-183; 191pp; English.

XX

CC The invention provides nucleic acid markers of prostate, breast and bladder cancer. The markers are indicators of malignant transformation of prostate, breast and bladder tissues and are diagnostic of the potential for metastatic spread of malignant prostate tumours. The nucleic acid can also be used as targets for therapeutic intervention in prostate cancer. The benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The markers may be used to design specific probes and primers, for the rapid

CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences from various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The probes and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX

SQ Sequence 2088 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 1 Other;

Query Match		100.0%;	Score 2087.6;	DB 3;	Length 2088;
Best Local Similarity		100.0%;	Pred No. 0;		
Matches 2088;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GACCTTAAATATATCAGGTGGCTAAATTCAGTGTATTAATTTACAAAATTTCTTCTTA	60		
DB	1	GACCTTAAATATATCAGGTGGCTAAATTCAGTGTATTAATTTACAAAATTTCTTCTTA	60		
QY	61	TTGCTACAGAGCTACAAATTCATTTACAGTAGGCCACCATGAGGCTCTTAAAGGAACC	120		
DB	61	TTGCTACAGAGCTACAAATTCATTTACAGTAGGCCACCATGAGGCTCTTAAAGGAACC	120		
QY	121	AGAAATATGAGGATATGCAATATTTATTTACATTTTACAGATCAGAAAATTTGAGGCACA	180		
DB	121	AGAAATATGAGGATATGCAATATTTATTTACATTTTACAGATCAGAAAATTTGAGGCACA	180		
QY	181	GATTAAGTAATCTCCCAAGGCTACAGGCATTCAGCTCCAGAAATCTGTCTTTACCAT	240		
DB	181	GATTAAGTAATCTCCCAAGGCTACAGGCATTCAGCTCCAGAAATCTGTCTTTACCAT	240		
QY	241	TCGTCTACAGGTATTTCCAAAAGAAAGAAAGTAAAAAGAGTCAAAAGGCAACAGAGT	300		
DB	241	TCGTCTACAGGTATTTCCAAAAGAAAGAAAGTAAAAAGAGTCAAAAGGCAACAGAGT	300		
QY	301	TCATGATTAATTCATAGAACAGTACACCATGCAATTTCTCACCCCTTTCAGACACACT	360		
DB	301	TCATGATTAATTCATAGAACAGTACACCATGCAATTTCTCACCCCTTTCAGACACACT	360		
QY	361	TGACCATGAAGGTTCTCTCAATGAATTTCTCATTTATTTTCAGAGCCATATTATTC	420		
DB	361	TGACCATGAAGGTTCTCTCAATGAATTTCTCATTTATTTTCAGAGCCATATTATTC	420		
QY	421	CATTGACTTTGAGTTTAACTCAGACCTAGGTCTGGAATGCTCTCTCTACTTATCCA	480		
DB	421	CATTGACTTTGAGTTTAACTCAGACCTAGGTCTGGAATGCTCTCTCTACTTATCCA	480		
QY	481	AAACTATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTCCAGAAA	540		
DB	481	AAACTATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTCCAGAAA	540		
QY	541	ATAAAAATGGTTGAAAGGCAATTTCTGCTACCAATGACTGTTTAAAGCCAGGCAAGTAA	600		
DB	541	ATAAAAATGGTTGAAAGGCAATTTCTGCTACCAATGACTGTTTAAAGCCAGGCAAGTAA	600		
QY	601	TGAACCATTTCCAACTCAATTTACTTATGAAAGAAATTTGATGATGAGGTTATTTTC	660		
DB	601	TGAACCATTTCCAACTCAATTTACTTATGAAAGAAATTTGATGATGAGGTTATTTTC	660		
QY	661	AAATTTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAATCATAGATTTATCT	720		
DB	661	AAATTTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAATCATAGATTTATCT	720		
QY	721	ATTATCTCAATTTAGTTTGTATTTATCTCTAGTGGGCCATTTAAATCTACCATGTGTT	780		
DB	721	ATTATCTCAATTTAGTTTGTATTTATCTCTAGTGGGCCATTTAAATCTACCATGTGTT	780		
QY	781	TCGTCTCTCCATTAGTCAATTAACCTAACTAGAGCAATTTAGTAAAGCCATGTGCCAGAT	840		
DB	781	TCGTCTCTCCATTAGTCAATTAACCTAACTAGAGCAATTTAGTAAAGCCATGTGCCAGAT	840		

QY	841	GCTCGCTAGGCACACAGAGGATAAAAACAATCTATTAGTATATACCACTAAATTTTCGCTT	900
DB	841	GCTCGCTAGGCACACAGAGGATAAAAACAATCTATTAGTATATACCACTAAATTTTCGCTT	900
QY	901	AGTAACTAGTGAATTTCAAGTCTGAGTCAAGAGTTGAGGAGACATTACAATGT	960
DB	901	AGTAACTAGTGAATTTCAAGTCTGAGTCAAGAGTTGAGGAGACATTACAATGT	960
QY	961	GTAATGAACCAAGGAAGTGAATTTGGATAGTGGGACCTAGTGTATTTATATTT	1020
DB	961	GTAATGAACCAAGGAAGTGAATTTGGATAGTGGGACCTAGTGTATTTATATATTT	1020
QY	1021	TAATTTGATTTCTGACTCTATCATTTGGCTCCAAACACAGATTTGTTTCTTTGGT	1080
DB	1021	TAATTTGATTTCTGACTCTATCATTTGGCTCCAAACACAGATTTGTTTCTTTGGT	1080
QY	1081	GTTTCTTTCATATGGGATCTTCTGTGCCAGCACAGTGCCTGACATAGAAAACAATC	1140
DB	1081	GTTTCTTTCATATGGGATCTTCTGTGCCAGCACAGTGCCTGACATAGAAAACAATC	1140
QY	1141	ATATTTTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1200
DB	1141	ATATTTTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1200
QY	1201	AGAAATATCCAGAGTAAAGTATGAGGCTCTGCTATTTATATGCGCTTAAATTAAGATT	1260
DB	1201	AGAAATATCCAGAGTAAAGTATGAGGCTCTGCTATTTATATGCGCTTAAATTAAGATT	1260
QY	1261	GTGAAAAGTTTAAAGACACTTAGTAGAGTGAATTTGAAATATATAGTAAACAATTTG	1320
DB	1261	GTGAAAAGTTTAAAGACACTTAGTAGAGTGAATTTGAAATATATAGTAAACAATTTG	1320
QY	1321	TGCTGGTCTTTTAAAGACATATTAATAGTAAATCTCCATCTCAAAAATAATG	1380
DB	1321	TGCTGGTCTTTTAAAGACATATTAATAGTAAATCTCCATCTCAAAAATAATG	1380
QY	1381	CATAAATATTTAAAGGAAATCAATCTCAGGCTTTCAATGTTTGTCTACTTTT	1440
DB	1381	CATAAATATTTAAAGGAAATCAATCTCAGGCTTTCAATGTTTGTCTACTTTT	1440
QY	1441	CATATATTTTACCATCTGCTGAGGAGTCTATCAAAAGGTTAAAGAAAGTGGAGGA	1500
DB	1441	CATATATTTTACCATCTGCTGAGGAGTCTATCAAAAGGTTAAAGAAAGTGGAGGA	1500
QY	1501	AACTCAGTAAAGATTTATTTAGTCTGTTTGCAAAAGTAAAGAAAGTCTCATCTCAA	1560
DB	1501	AACTCAGTAAAGATTTATTTAGTCTGTTTGCAAAAGTAAAGAAAGTCTCATCTCAA	1560
QY	1561	CTTTATGAGCAGGAAGAGGAGGCTGTTTGAAGACCATTTTCTAGCAGAACCATAT	1620
DB	1561	CTTTATGAGCAGGAAGAGGAGGCTGTTTGAAGACCATTTTCTAGCAGAACCATAT	1620
QY	1621	TTTAGACACTTCCCTGCAATTAATCTGCAAAACAATATGTTTGCAAAAGTGGTATCAAC	1680
DB	1621	TTTAGACACTTCCCTGCAATTAATCTGCAAAACAATATGTTTGCAAAAGTGGTATCAAC	1680
QY	1681	CTCCAAACAGACACATTTAGGAGTTAAATATTTTTCATCAAAAGTGGTATTTCTTTA	1740
DB	1681	CTCCAAACAGACACATTTAGGAGTTAAATATTTTTCATCAAAAGTGGTATTTCTTTA	1740
QY	1741	ACGCTAGAGATTGCTACAAATCTTCTGAGGCTCAATGCTTCAAGGCTTCAAGAGAT	1800
DB	1741	ACGCTAGAGATTGCTACAAATCTTCTGAGGCTCAATGCTTCAAGGCTTCAAGAGAT	1800
QY	1801	TTCTCCCTGTTATAGCAGCAAGCAAAATTTAGCCTTTCTCTCAAACTTCACTAATGA	1860
DB	1801	TTCTCCCTGTTATAGCAGCAAGCAAAATTTAGCCTTTCTCTCAAACTTCACTAATGA	1860
QY	1861	TCACATTTCTTCCAAAGGAATCTCTAGAGACCAATGCCCCGAGTTAAGAACATCAAAA	1920
DB	1861	TCACATTTCTTCCAAAGGAATCTCTAGAGACCAATGCCCCGAGTTAAGAACATCAAAA	1920
QY	1921	CTAACCATCTGAAAGAACTTCCCAAGTGTAAAGCTCTCTGCTGCAAGCAACATAAAAA	1980

```
Db 1921 CTAACCATCTGAAGAACTTCCCAAGTGAAGACTCTGCTGCACACACATAAAAA 1980
Qy 1981 AAGAGAGAAGAAATCAATAGACACAAATAAATAAGGGGATATCACCACCGATCC 2040
Db 1981 AAGAGAGAAGAAATCAATAGACACAAATAAATAAGGGGATATCACCACCGATCC 2040
Qy 2041 CACAGAAATACAACTACCATCGAGAACTACAAACACCTCTAGGC 2088
Db 2041 CACAGAAATACAACTACCATCGAGAACTACAAACACCTCTAGGC 2088

RESULT 2
ID AAS04000 standard; cDNA; 2087 BP.
AC AAS04000;
XX
XX 29-AUG-2001 (first entry)
XX
XX Biomarker UC band 28 #2, used in diagnosis and prognosis of cancer.
DE
XX
XX Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 99..506
FT CDS /*tag= a
FT /product= "Prostate cancer marker protein"
XX
XX US6218529-B1.
XX
XX 17-APR-2001.
XX
XX 12-JUN-1998; 98US-00097199.
XX
XX 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.
XX
XX (UROC-) UROCOR INC.
XX
XX An G, O'hara SM, Ralph D, Veltri R;
XX
XX WPI; 2001-289849/30.
XX
XX P-PSDB; AAU02174.
XX
XX New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
XX Claim 2; Col 117-121; 78pp; English.
XX
XX The sequence represents nucleic acid biomarker UC band 28 #2, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that
CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX
XX Sequence 2087 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 0 Other;
XX
XX Query Match 99.4%; Score 2076; DB 4; Length 2087;
XX Best Local Similarity 100.0%; Pred. No. 0;
```

```
Matches 2087; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1 GACCTTAAATATATATGAGTGGCTAAATTTGATGATATAATAATTTACAAAATTTATCTTCTTA 60
Db 1 GACCTTAAATATATGAGTGGCTAAATTTGATGATATAATAATTTACAAAATTTATCTTCTTA 60
Qy 61 TTGCTACAGAGCTACAATTTCAATTTTACAGTAGGCGCCACCATGAGGCGCTTCTTTAAGGAACC 120
Db 61 TTGCTACAGAGCTACAATTTCAATTTTACAGTAGGCGCCACCATGAGGCGCTTCTTTAAGGAACC 120
Qy 121 AGAAATATGAGGATATGCACAATATTTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
Db 121 AGAAATATGAGGATATGCACAATATTTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
Qy 181 GATTAAGTAACTTCCCAAGGCTACAGGCAATTTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
Db 181 GATTAAGTAACTTCCCAAGGCTACAGGCAATTTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
Qy 241 TCTGCTACAAGGTATTTTGGAAAAAGAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 300
Db 241 TCTGCTACAAGGTATTTTGGAAAAAGAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 300
Qy 301 TCATTGATTTATTCATAGAACAGTCAACACCATGCAATTTCTCACCCCTTGCAGACACACT 360
Db 301 TCATTGATTTATTCATAGAACAGTCAACACCATGCAATTTCTCACCCCTTGCAGACACACT 360
Qy 361 TGACCATGAAAGGTTCTCAATGAATGTTCCTCATATTTCTTTCAAGAGCCATATTATCCA 420
Db 361 TGACCATGAAAGGTTCTCAATGAATGTTCCTCATATTTCTTTCAAGAGCCATATTATCCA 420
Qy 421 CATTCACCTTTGCGAGTTAACTGAGCCCTAGGTCTGGAATGCTGCTCTCTCTATTATCCA 480
Db 421 CATTCACCTTTGCGAGTTAACTGAGCCCTAGGTCTGGAATGCTGCTCTCTCTATTATCCA 480
Qy 481 AAACCTATACATCCACAGATCATATAAACTCTCAGCCCTGCTCGAAAGCCCTTTCCAGAAAA 540
Db 481 AAACCTATACATCCACAGATCATATAAACTCTCAGCCCTGCTCGAAAGCCCTTTCCAGAAAA 540
Qy 541 ATAAAAATGGTTGAAAGGCAATCTGCTCAATGACCTGTTTAAAGCCAGCAAGTAAC 600
Db 541 ATAAAAATGGTTGAAAGGCAATCTGCTCAATGACCTGTTTAAAGCCAGCAAGTAAC 600
Qy 601 TGAACCATTCNACTTCAATTTTACATGAAAGAAATTTGATGATGATGAGGAGTTATTC 660
Db 601 TGAACCATTCNACTTCAATTTTACATGAAAGAAATTTGATGATGATGAGGAGTTATTC 660
Qy 661 AATTCCTAAAATACAAACCAGTGGTGAATCTTCTCAATCTTGAACCTCATAGATTATATCT 720
Db 661 AATTCCTAAAATACAAACCAGTGGTGAATCTTCTCAATCTTGAACCTCATAGATTATATCT 720
Qy 721 ATATATCTCAATTTAGTGTGTTTATCTAGTGGGCCATTAATAAACCTACCAATGTGT 780
Db 721 ATATATCTCAATTTAGTGTGTTTATCTAGTGGGCCATTAATAAACCTACCAATGTGT 780
Qy 781 TCTGTCTCTCCATAGTCAATTAACCTAACTAAGAGCAATTTAGTAAAGCCATGTGCAGAT 840
Db 781 TCTGTCTCTCCATAGTCAATTAACCTAACTAAGAGCAATTTAGTAAAGCCATGTGCAGAT 840
Qy 841 GCTCCGCTAGGCACAGAGGATAAAAACAATCTATAGTATACCACTAAATTTTCGCTT 900
Db 841 GCTCCGCTAGGCACAGAGGATAAAAACAATCTATAGTATACCACTAAATTTTCGCTT 900
Qy 901 AGTAACTAGTCAAAATGTTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Db 901 AGTAACTAGTCAAAATGTTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTACAATGT 960
Qy 961 GTAATGGAAACCAAGGAAAGTGAACCTTTGATAGTGGGAGCTAGTGTATTTATATTT 1020
Db 961 GTAATGGAAACCAAGGAAAGTGAACCTTTGATAGTGGGAGCTAGTGTATTTATATTT 1020
Qy 1021 TAAATGATTTCTGACTCTATCATTTGGCCCTCCAAACACAGATTTGTTTTCTTGGTTTT 1080
Db 1021 TAAATGATTTCTGACTCTATCATTTGGCCCTCCAAACACAGATTTGTTTTCTTGGTTTT 1080
```

```
QY 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCACAGTGCCTGTGACATAGAGAAAATC 1140
Db 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCACAGTGCCTGTGACATAGAGAAAATC 1140
QY 1141 AATATTTGCTGAATTAATGATTAATAAATACAGAGAACTTTCCATCTCTGTTTGGATCTAT 1200
Db 1141 AATATTTGCTGAATTAATGATTAATAAATACAGAGAACTTTCCATCTCTGTTTGGATCTAT 1200
QY 1201 AGAACATCCAGAGTAAGTATGATGAGGCGCTCTGCATTTATATGCGCTTAATTAAGATTAAT 1260
Db 1201 AGAACATCCAGAGTAAGTATGATGAGGCGCTCTGCATTTATATGCGCTTAATTAAGATTAAT 1260
QY 1261 GTGAGAAAGTTTAAAGACACCTTAGTAGAGTATTTGAAATATATAGTAAACACATTTGGAAA 1320
Db 1261 GTGAGAAAGTTTAAAGACACCTTAGTAGAGTATTTGAAATATATAGTAAACACATTTGGAAA 1320
QY 1321 TGGTGGTGTGTTTAAAGAGATATTAATAGATAATATGAAATCTCCATCTCAAAAATAATG 1380
Db 1321 TGGTGGTGTGTTTAAAGAGATATTAATAGATAATATGAAATCTCCATCTCAAAAATAATG 1380
QY 1381 CATAACTATTAAAGGAAATCAATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1440
Db 1381 CATAACTATTAAAGGAAATCAATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1440
QY 1441 CATATATTTTACCCTCTGCTGAAGCCAGTCAATCAAGGGTAAAGAAAGATGGGAGA 1500
Db 1441 CATATATTTTACCCTCTGCTGAAGCCAGTCAATCAAGGGTAAAGAAAGATGGGAGA 1500
QY 1501 AAATCAGTAGAATATATATAGTCTGTTTGCAGAGTAGAAAGATCTCATCACTCAA 1560
Db 1501 AAATCAGTAGAATATATATAGTCTGTTTGCAGAGTAGAAAGATCTCATCACTCAA 1560
QY 1561 CTTATGACGAGAGAGGAGGCTGTTTGAGAACCTTTACTTAGCAGAACCATAT 1620
Db 1561 CTTATGACGAGAGAGGAGGCTGTTTGAGAACCTTTACTTAGCAGAACCATAT 1620
QY 1621 TTTAGACACTTCCCTGCAATTAAGTCAACAAATATGTTTGCAAACTGTTTGTGATCAAC 1680
Db 1621 TTTAGACACTTCCCTGCAATTAAGTCAACAAATATGTTTGCAAACTGTTTGTGATCAAC 1679
QY 1681 CTCCAAACGACACATTCAGAGTTAAATATTTTTCATCAAAATTTGGATTTTCCCTTA 1740
Db 1680 CTCCAAACGACACATTCAGAGTTAAATATTTTTCATCAAAATTTGGATTTTCCCTTA 1739
QY 1741 ACGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAGAGAT 1800
Db 1740 ACGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAGAGAT 1799
QY 1801 TTCTCCCTGTTTATAAGCAGCAAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATGA 1860
Db 1800 TTCTCCCTGTTTATAAGCAGCAAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATGA 1859
QY 1861 TCACATTTCTTCCAAAGAACTCTAGAGACCAATGCCCCGGAGTTTAAAGACATCAAAA 1920
Db 1860 TCACATTTCTTCCAAAGAACTCTAGAGACCAATGCCCCGGAGTTTAAAGACATCAAAA 1919
QY 1921 CTAACCATCTGAAGAACTTCCCAAGTGTAAAGCTCTGCTGCGACGACAAACATAAAAA 1980
Db 1920 CTAACCATCTGAAGAACTTCCCAAGTGTAAAGCTCTGCTGCGACGACAAACATAAAAA 1979
QY 1981 AAGAGAGAAGAAATCAATATAGACACAATAAAAATGATAAAGGGGATATCAACCCGATCC 2040
Db 1980 AAGAGAGAAGAAATCAATATAGACACAATAAAAATGATAAAGGGGATATCAACCCGATCC 2039
QY 2041 CACAGAAATACAACTACCATCAGAGATATCTACAAAACACCTCTTACGC 2088
Db 2040 CACAGAAATACAACTACCATCAGAGATATCTACAAAACACCTCTTACGC 2087
```

```
XX AAZ87584;  
XX 19-APR-2000 (first entry)  
XX Prostate disease marker UC Band #28 splice variant.  
XX Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;  
XX benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;  
XX diagnosis; PCR primer; ss.  
XX Homo sapiens.  
XX WO9964631-A1.  
XX 16-DEC-1999.  
XX 11-JUN-1999; 99WO-US013151.  
XX 12-JUN-1998; 98US-00097199.  
XX (UROC-) UROCOR INC.  
XX An G, O'hara SM, Ralph D, Veltri RW;  
XX WPI; 2000-116557/10.  
XX P-PSDB; AAY59296.  
XX Novel RNA biomarkers for diagnosis, prognosis and management of prostate,  
XX breast and bladder cancer.  
XX Claim 1; Page 184-186; 191pp; English.  
XX The invention provides nucleic acid markers of prostate, breast and  
XX bladder cancer. The markers are indicators of malignant transformation of  
XX prostate, breast and bladder tissues and are diagnostic of the potential  
XX for metastatic spread of malignant prostate tumours. The nucleic acid can  
XX also be used as targets for therapeutic intervention in prostate cancer,  
XX benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The  
XX markers may be used to design specific probes and primers, for the rapid  
XX analysis of prostate, bladder or breast biopsy samples. The probes and  
XX primers may also be used for in situ hybridization or in situ PCR  
XX detection and diagnosis. They may also be used to identify and isolate  
XX full length gene sequences from various DNA libraries. Antibodies against  
XX the polypeptide products of the markers can be used to treat prostate  
XX cancer, bladder cancer or breast cancer. The encoded proteins may be used  
XX to detect antibodies. The proteins and antibodies can be used in  
XX immunodetection methods for detecting or quantifying the cancers, and for  
XX clinical diagnosis of these cancers. The antibodies may also be used for  
XX radioimaging to quantify and localize the encoded proteins  
XX  
SQ Sequence 2506 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 1 Other;
```

Query Match 93.9%; Score 1961.6; DB 3; Length 2506;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1968; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
QY 1 GACCTTAAATATATCGAGGTGGCTAAATTTAGATAGTATATAATTTACAAAATATTCTTCTA 60  
Db 1 GACCTTAAATATATCGAGGTGGCTAAATTTAGATAGTATATAATTTACAAAATATTCTTCTA 60  
QY 61 TTGCTACAGACTACAATTTCAATTTACAGTAGGCCACCATGAGGGCCCTTCTTAAGGAACC 120  
Db 61 TTGCTACAGACTACAATTTCAATTTACAGTAGGCCACCATGAGGGCCCTTCTTAAGGAACC 120  
QY 121 AGAAATATGAGGATATGCACAAATATTATTCATTTACAGATCAGAAAATTTAGGCGACA 180  
Db 121 AGAAATATGAGGATATGCACAAATATTATTCATTTACAGATCAGAAAATTTAGGCGACA 180  
QY 181 GATTAAAGTAACTTCCCAAGGCTACAGGCATTTAGCTCCAGAACTGTGCTCTTACCAT 240  
Db 181 GATTAAAGTAACTTCCCAAGGCTACAGGCATTTAGCTCCAGAACTGTGCTCTTACCAT 240
```



```
QY 241 TCTGCTACAGGTATTTTCGAAAAAGAAAAAGTAAAAAGTCAAAAGGCAACAGAGT 300
DB 241 TCTGCTACAGGTATTTTCGAAAAAGAAAAAGTAAAAAGTCAAAAGGCAACAGAGT 300
QY 301 TCATTGATTATTCATAGAACAGTCAACATGCAATCTCACACCCCTTTCGAGACACACT 360
DB 301 TCATTGATTATTCATAGAACAGTCAACATGCAATCTCACACCCCTTTCGAGACACACT 360
QY 361 TGACCATGAAGGTTCCTCAATGAATGTTCCCTCATTTATCTTCAGAGCCATATTATTC 420
DB 361 TGACCATGAAGGTTCCTCAATGAATGTTCCCTCATTTATCTTCAGAGCCATATTATTC 420
QY 421 CATTTGACTTTCAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTCTCTACTTATCCA 480
DB 421 CATTTGACTTTCAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTCTCTACTTATCCA 480
QY 481 AAATACTATACATCACAGATCATATAAATCTCTAGCCCTGCTGCAAGCCCTTTCAGAAAA 540
DB 481 AAATACTATACATCACAGATCATATAAATCTCTAGCCCTGCTGCAAGCCCTTTCAGAAAA 540
QY 541 ATAAAAATGGTTGAAAGGCAATTTCTGTACCAATGACTGTTTAAAGCCAGCAAGTAAC 600
DB 541 ATAAAAATGGTTGAAAGGCAATTTCTGTACCAATGACTGTTTAAAGCCAGCAAGTAAC 600
QY 601 TGAACCAATTCCTCAATTTACTTATGAAAGAAATTTGATGATGTAGAGGTTATTTTC 660
DB 601 TGAACCAATTCCTCAATTTACTTATGAAAGAAATTTGATGATGTAGAGGTTATTTTC 660
QY 661 AATTCTAAAAATACAAAACCAATGTTGATCTTCTCAATCTTGAATCATAGATTATTTATCT 720
DB 661 AATTCTAAAAATACAAAACCAATGTTGATCTTCTCAATCTTGAATCATAGATTATTTATCT 720
QY 721 ATTATCTCAATTTAGTTTGTATTTATCTCTAGTGGGCCATTTAAAAACTACCAATGTGTT 780
DB 721 ATTATCTCAATTTAGTTTGTATTTATCTCTAGTGGGCCATTTAAAAACTACCAATGTGTT 780
QY 781 TCTGCTCTCCATTAGTCAATAAATCTAACTAAGAGCAATTTAGTAAAGCCATGTCCAGAT 840
DB 781 TCTGCTCTCCATTAGTCAATAAATCTAACTAAGAGCAATTTAGTAAAGCCATGTCCAGAT 840
QY 841 GCTCCGCTAGGCACAGAGGATAAAAAACAATCTTATAGTATACCACATAATTTTCGCTT 900
DB 841 GCTCCGCTAGGCACAGAGGATAAAAAACAATCTTATAGTATACCACATAATTTTCGCTT 900
QY 901 AGTAACCTAGTGAATGTTCAAGTCAATGCTGAGTCAAGAGTTTGAGGAGACATTTACAATGT 960
DB 901 AGTAACCTAGTGAATGTTCAAGTCAATGCTGAGTCAAGAGTTTGAGGAGACATTTACAATGT 960
QY 961 GTAAATGGAACCAAGGAAAGTGAATCTTTGGATAAGTGGGGACTAGTGTATTATATATT 1020
DB 961 GTAAATGGAACCAAGGAAAGTGAATCTTTGGATAAGTGGGGACTAGTGTATTATATATT 1020
QY 1021 TAATTGATTTCTGACTCTATCATTTGGCCCTCCAAACACAGATGTTGTTTTCTTTGGTTTT 1080
DB 1021 TAATTGATTTCTGACTCTATCATTTGGCCCTCCAAACACAGATGTTGTTTTCTTTGGTTTT 1080
QY 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGACACAGTGCCTGACACATAGAAAAACAATC 1140
DB 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGACACAGTGCCTGACACATAGAAAAACAATC 1140
QY 1141 AATATTTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCATTTCTGTTTGGATCTAT 1200
DB 1141 AATATTTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCATTTCTGTTTGGATCTAT 1200
QY 1201 AGAACATCCAGAGTATGATGAGGCGCTCTGATTTATATGCTTTAAATTAAGATTAT 1260
DB 1201 AGAACATCCAGAGTATGATGAGGCGCTCTGATTTATATGCTTTAAATTAAGATTAT 1260
QY 1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGAATTTTGAATATATAGTAAACACTTGGAAA 1320
DB 1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGAATTTTGAATATATAGTAAACACTTGGAAA 1320
QY 1321 TGTGTGTGCTTTTAAAAAGATATTAATAGATAATATGAAAAATCTCCATCTCAAAAAATAATG 1380
```

```
DB 1321 TGTGTGTGCTTTTAAAAAGATATTAATAGATAATATGAAAAATCTCCATCTCAAAAAATAATG 1380
QY 1381 CATAAACTATTATTAAGGAAAAATCATCTCTCAGGCTTTCAATGTTTGTTCATTTACTTTTT 1440
DB 1381 CATAAACTATTATTAAGGAAAAATCATCTCTCAGGCTTTCAATGTTTGTTCATTTACTTTTT 1440
QY 1441 CATATATTTTACCATCTGCTGAAGGAGTGTATCAAAAGGGTAAAGAAAGATGGGAGA 1500
DB 1441 CATATATTTTACCATCTGCTGAAGGAGTGTATCAAAAGGGTAAAGAAAGATGGGAGA 1500
QY 1501 AAACCTCAGTAAAGATATATATTAGTCTGTTTGCAAAAGTAAAGAAAGATTTCTCATCTCAA 1560
DB 1501 AAACCTCAGTAAAGATATATATTAGTCTGTTTGCAAAAGTAAAGAAAGATTTCTCATCTCAA 1560
QY 1561 CCTTATGAGCAGGAAGAGGAGGCTGTTTGAAGAACCATTTTACTTTAGCAGAACCATAT 1620
DB 1561 CCTTATGAGCAGGAAGAGGAGGCTGTTTGAAGAACCATTTTACTTTAGCAGAACCATAT 1620
QY 1621 TTTAGACACTTCCCTGCAATTAATCTGCAACAAATATGTTTGCAAAAGTAAAGAAAGATTTCTCATCTCAA 1680
DB 1621 TTTAGACACTTCCCTGCAATTAATCTGCAACAAATATGTTTGCAAAAGTAAAGAAAGATTTCTCATCTCAA 1680
QY 1681 CTCCAAACAGACACATTTAGGAGTTAAATATTTTTCATCAAAACATTTGATTTTCTTTA 1740
DB 1681 CTCCAAACAGACACATTTAGGAGTTAAATATTTTTCATCAAAACATTTGATTTTCTTTA 1740
QY 1741 ACGCTAGAGATTGCTACAAATCTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT 1800
DB 1741 ACGCTAGAGATTGCTACAAATCTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT 1800
QY 1801 TTCTCCCTGTTTAAAGCAGCAAGCAATTAAGCCATTTCTCTCAAACTTCACTAAATGA 1860
DB 1801 TTCTCCCTGTTTAAAGCAGCAAGCAATTAAGCCATTTCTCTCTCAAACTTCACTAAATGA 1860
QY 1861 TCACATCTTCTCCAAAGGAACTCTAGAGACCAAAATGCCGAGTTAAGAACATCAAAA 1920
DB 1861 TCACATCTTCTCCAAAGGAACTCTAGAGACCAAAATGCCGAGTTAAGAACATCAAAA 1920
QY 1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAAGACTCTGCTGACGACGACAAACATATAA 1978
DB 1921 CTAACCATCTGAAGAAACTTCCCAAGTGTAAAGACTCTGCTGACGACGACAAACATATAA 1978
```

RESULT 4

AAS04001

ID AAS04001 standard; cDNA; 2505 BP.

XX AC AAS04001;

XX DT 29-AUG-2001 (first entry)

XX DE Biomarker UC band 28 #3, used in diagnosis and prognosis of cancer.

XX KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;

XX KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.

XX OS Homo sapiens.

XX FH Key

XX FT CDS

XX FT Location/Qualifiers

XX FT 99..506

XX FT /*tag= a

XX FT /product= "Prostate cancer marker protein"

XX PN US6218529-B1.

XX PD 17-APR-2001.

XX PF 12-JUN-1998; 98US-00097199.

XX PR 31-JUL-1995; 95US-0001655P.

XX PR 11-JAN-1996; 96US-0013611P.

XX PR 31-JUL-1996; 96US-00692787.

XX (UROC-) UROCOR INC.
XX An G, O'hara SM, Ralph D, Veltri R;
XX WPI; 2001-289849/30.
XX DR P-PSDB; AAU02175.
XX
PT New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
XX and bladder cancer.
XX
PS Claim 2; Col 121-125; 78pp; English.
XX
CC The sequence represents nucleic acid biomarker UC band 28 #3, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that
CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ Sequence 2505 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 0 Other;

Query Match 93.4%; Score 1950; DB 4; Length 2505;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 GACCTTAATATATATGAGTGGCTGAATGATGATTAATTAATTTACAAAATTAATCTCTTA 60
DB 1 GACCTTAATATATATGAGTGGCTGAATGATGATTAATTAATTTACAAAATTAATCTCTTA 60
QY 61 TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCATGAGGCGCTCTTTAAGGAACC 120
DB 61 TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCATGAGGCGCTCTTTAAGGAACC 120
QY 121 AGAAATATGAGGATATGCAATATTAATTTACAGATCAAGAAATTTGAGGCACA 180
DB 121 AGAAATATGAGGATATGCAATATTAATTTACAGATCAAGAAATTTGAGGCACA 180
QY 181 GATTAAAGTAACCTCCAGGCTACAGGCAATCTAGCTCCAGAACTGTGCTTTACCAT 240
DB 181 GATTAAAGTAACCTCCAGGCTACAGGCAATCTAGCTCCAGAACTGTGCTTTACCAT 240
QY 241 TCTGCTACAAAGGTATTTTCAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 300
DB 241 TCTGCTACAAAGGTATTTTCAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 300
QY 301 TCATTGATTATTTCCATAGAACAGTCAACCATGCAATTTCTCACCCCTTGCAGACACACT 360
DB 301 TCATTGATTATTTCCATAGAACAGTCAACCATGCAATTTCTCACCCCTTGCAGACACACT 360
QY 361 TGACCATGAAAGGTTCTCAATGAATGTTCTCTATTATCTTCAGAGGCCATATTATCA 420
DB 361 TGACCATGAAAGGTTCTCAATGAATGTTCTCTATTATCTTCAGAGGCCATATTATCA 420
QY 421 CATTTGACTTTGAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTCTCTACTTTATCCA 480
DB 421 CATTTGACTTTGAGTTAACTCAGACCCCTAGGTTGGAATGCTGCTCTCTACTTTATCCA 480
QY 481 AAACTATACATCCAGAGTATATAAATCTTCAGCCCTGCTGCAAGCCCTTTCCAGAAAA 540
DB 481 AAACTATACATCCAGAGTATATAAATCTTCAGCCCTGCTGCAAGCCCTTTCCAGAAAA 540
QY 541 ATAAAAATGGTTGAAAGGCAATCTGCTACCAATGACTGTTTAAGCCCGCAGCAAGTAAC 600
DB 541 ATAAAAATGGTTGAAAGGCAATCTGCTACCAATGACTGTTTAAGCCCGCAGCAAGTAAC 600

QY 601 TGAACCAATTCAACTTCAATTTTACTTATGAAAAAGAAATTTGATGATGTAGGAGGTTATTTTC 660
DB 601 TGAACCAATTCAACTTCAATTTTACTTATGAAAAAGAAATTTGATGATGTAGGAGGTTATTTTC 660
QY 661 AATCTTAAAAATACAAACCCCATGTTGATCTTCTCAATCTTGAACCTCATAGATTATTTATCT 720
DB 661 AATCTTAAAAATACAAACCCCATGTTGATCTTCTCAATCTTGAACCTCATAGATTATTTATCT 720
QY 721 ATTATCTCAATTTAGTTTGTATTATCTAGTGGGCCATTAAAAAAGTACCAATGTTT 780
DB 721 ATTATCTCAATTTAGTTTGTATTATCTAGTGGGCCATTAAAAAAGTACCAATGTTT 780
QY 781 TCTGTCTCTCCATCTAGTCAATAAATACTAACTAAAGAGCAATTTAGTAAAGCCATGTGCCAGAT 840
DB 781 TCTGTCTCTCCATCTAGTCAATAAATACTAACTAAAGAGCAATTTAGTAAAGCCATGTGCCAGAT 840
QY 841 GCTCGCTAGGACCCAGAGGAGTAAAAACAATCTTATAGTATACCACTAATTTTCGCTT 900
DB 841 GCTCGCTAGGACCCAGAGGAGTAAAAACAATCTTATAGTATACCACTAATTTTCGCTT 900
QY 901 AGTAACTAGTGAATGTTTCAAGTCATGCTCAGTCAAGAGTTGAGGAGACATTACAATGT 960
DB 901 AGTAACTAGTGAATGTTTCAAGTCATGCTCAGTCAAGAGTTGAGGAGACATTACAATGT 960
QY 961 GTAATGGAACCAAGGAAAGTGAAACTTTTGATTAAGTGGGACCTAGTCTATTTATATATT 1020
DB 961 GTAATGGAACCAAGGAAAGTGAAACTTTTGATTAAGTGGGACCTAGTCTATTTATATATT 1020
QY 1021 TAAATGATTTCTGACTCTATCTATCAITGGCCCTCAAAACACAGATTGTGTTTCTTTGGTTTT 1080
DB 1021 TAAATGATTTCTGACTCTATCTATCAITGGCCCTCAAAACACAGATTGTGTTTCTTTGGTTTT 1080
QY 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCAGCAGTGCCTGCACACATAGAAAAACAATC 1140
DB 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCAGCAGTGCCTGCACACATAGAAAAACAATC 1140
QY 1141 AATATTTCTCTCAATTAATTAATTAATAAATCAGAGAACTTTTCCCATCTCTGTTTGGATCTAT 1200
DB 1141 AATATTTCTCTCAATTAATTAATTAATAAATCAGAGAACTTTTCCCATCTCTGTTTGGATCTAT 1200
QY 1201 AGAACATCCAGAGTAAGTGAAGGCGCTCTGCAATTTATATGCGCTTAAAATTAAGATTAT 1260
DB 1201 AGAACATCCAGAGTAAGTGAAGGCGCTCTGCAATTTATATGCGCTTAAAATTAAGATTAT 1260
QY 1261 GTGAGAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAATAATAGTAAACACTTGGAAA 1320
DB 1261 GTGAGAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAATAATAGTAAACACTTGGAAA 1320
QY 1321 TGGTGGTCTTTAAAAAGATATTAATAGATAATATGAAAACTCCATCTCAAAAAATAATG 1380
DB 1321 TGGTGGTCTTTAAAAAGATATTAATAGATAATATGAAAACTCCATCTCAAAAAATAATG 1380
QY 1381 CATAAACTATTTAAAGGAAAAATCACTCAGGCTTTCAATGTTTGTCTACTTTT 1440
DB 1381 CATAAACTATTTAAAGGAAAAATCACTCAGGCTTTCAATGTTTGTCTACTTTT 1440
QY 1441 CATATATTTTACCCTGCTGAAGGAGTCAATATCAAGGCTTAAAGAAAGTGGAGGA 1500
DB 1441 CATATATTTTACCCTGCTGAAGGAGTCAATATCAAGGCTTAAAGAAAGTGGAGGA 1500
QY 1501 AAATCAGTAAGAAATTAATTTAGTCTGTTTGAAGTAGAAAAAGATTTCTCATCTCAAC 1560
DB 1501 AAATCAGTAAGAAATTAATTTAGTCTGTTTGAAGTAGAAAAAGATTTCTCATCTCAAC 1560
QY 1561 CCTTATAGCAGGAGGAGGAGGCTGTTTGAAGCCATTTTCTTAGCAGAACCAATAT 1620
DB 1561 CCTTATAGCAGGAGGAGGAGGCTGTTTGAAGCCATTTTCTTAGCAGAACCAATAT 1620
QY 1621 TTTAGACACTTCCCTGCAATTAAGTGCACAAACATATGTTTGCACAACTGTTTGTGATCAAC 1680
DB 1621 TTTAGACACTTCCCTGCAATTAAGTGCACAAACATATGTTTGCACAACTGTTTGTGATCAAC 1680
QY 1621 TTTAGACACTTCCCTGCAATTAAGTGCACAAACATATGTTTGCACAACTGTTTGTGATCAAC 1680
DB 1621 TTTAGACACTTCCCTGCAATTAAGTGCACAAACATATGTTTGCACAACTGTTTGTGATCAAC 1680

QY 1681 CTCACACAGCACACATTCAGGAGTTAAATATATTTTTCATCAACATTTGGATTTTCCTTA 1740
Db CTTCCACACAGCACACATTCAGGAGTTAAATATATTTTTCATCAACATTTGGATTTTCCTTA 1739
QY 1741 ACCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAAGGCTTAAGAGAGAT 1800
Db 1740 ACCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAAGGCTTAAGAGAGAT 1799
QY 1801 TTCTCCTCTTTATPAGCAGACAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATGA 1860
Db 1800 TTCTCCTCTTTATPAGCAGACAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATGA 1859
QY 1861 TCACATTTCTTCCAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGACATCAAAA 1920
Db 1860 TCACATTTCTTCCAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGACATCAAAA 1919
QY 1921 CTAAACCATCTGAAGAACTTCCCAAGTGTAAAGACTCTGCTGCACGACACATATAA 1978
Db 1920 CTAAACCATCTGAAGAACTTCCCAAGTGTAAAGACTCTGCTGCACGACATTAACATTAACGAGA 1977

RESULT 5

AAV16883

ID AAV16883 standard; DNA; 757 BP.

XX AAV16883;

AC AAV16883;

DT 07-AUG-1998 (first entry)

DE Human prostate cancer marker UC Band #28.

KW Prostate cancer; human; marker; diagnosis; treatment; probe; ss.

OS Homo sapiens.

PN WO9804689-A1.

XX 05-FEB-1998.

PF 31-JUL-1996; 96WO-US012516.

XX 31-JUL-1996; 96WO-US012516.

PR (UROC-) UROCOR INC.

PA Veltri R, Ohara SM, An G, Ralph D;

XX WPI; 1998-130681/12.

DR Human prostate cancer marker - useful for detection and treatment of

PT human prostate cancer.

XX Claim 1; Page 133-134; 229pp; English.

XX This represents a marker sequence for human prostate cancer. Isolated

CC nucleic acid segments shown in AAV16881 to AAV16885, AAV16890 to

CC AAV16903, AAV26351 and AAV26352 which can act as human prostate cancer

CC markers are provided in the specification. It also provides methods for

CC identifying markers for human prostate cancer and for detection of

CC prostate cancer cells. The markers can be identified by amplifying human

CC prostate RNA to provide nucleic acid amplification products, separating

CC the products and identifying those RNA that are differentially expressed

CC between human prostate cancers versus normal or benign human prostate.

CC Prostate cancer cells in a sample can be detected by detecting a nucleic

CC acid in a sample, the nucleic acid being a prostate cancer marker.

CC Primers and probes derived from this marker can be used for the detection

CC of prostate cancer cells in a sample. Antibodies against the protein

CC encoded by the marker nucleic acid fragments, inhibitors of the protein

CC and oligonucleotides antisense to the markers can be used in the

CC treatment of prostate cancer. The antibodies can also be used for the

CC diagnosis of human prostate cancer

XX Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

SQ

Query Match 13.2%; Score 275.4; DB 2; Length 757;
Best Local Similarity 96.2%; Pred. NO. 9.3e-53;
Matches 282; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1686 ACAACGACACATTCAGGAGTTAAATATATTTTTCATCAACATTTGGATTTTCCTTAACGCT 1745
Db 1 ACAACGACACATTCAGGAGTTAAATATATTTTTCATCAACATTTGGATTTTCCTTAACGCT 60
QY 1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAAGGCTTAAGAGAGATTTCTC 1805
Db 61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAAGGCTTAAGAGAGATTTCTC 120
QY 1806 CCTGTTTAAAGCAGACAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATGATCACA 1865
Db 121 CCTGTTTAAAGCAGACAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATGATCACA 180
QY 1866 TTCTTTTCCAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGACATCAAAAATTAAC 1925
Db 181 TTCTTTTCCAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGACATCAAAAATTAAC 240
QY 1926 CATCTGAAGAACTTCCCAAGTGTAAAGACTCTGCTGCACGACACATATAA 1978
Db 241 CATCTGAAGAACTTCCCAAGTGTAAAGACTCTGCTGCATTAACATTAACGAGA 293

RESULT 6

AAAX26018

ID AAX26018 standard; DNA; 757 BP.

XX AAX26018;

AC AAX26018;

DT 20-MAY-1999 (first entry)

DE Prostate disease marker gene fragment UC Band #28.

KW Prostate cancer; benign prostatic hyperplasia; marker gene; tumour;

KW differentiation; Reverse Transcription Polymerase Chain Reaction;

KW diagnostic; progression; cancer; metastasis; human; RT-PCR; ss.

XX Homo sapiens.

XX US5882864-A.

PN 16-MAR-1999.

XX 31-JUL-1996; 96US-00692787.

PF 31-JUL-1995; 95US-0001655P.

PR (UROC-) UROCOR INC.

PA Veltri R, Ralph D, An G, O'hara SM;

XX WPI; 1999-214055/18.

DR Diagnosing prostate cancer and benign prostatic hyperplasia cells - using

PT oligonucleotide probes specific for marker genes associated with tumor

PT differentiation and progression in Reverse Transcription Polymerase Chain

PT Reaction analysis.

XX Claim 1; Col 71-72; 74pp; English.

XX The invention relates to methods for diagnosing prostate cancer or benign

CC prostatic hyperplasia cells in a biological sample. The method uses

CC oligonucleotide probes specific for marker genes associated with tumor

CC differentiation and progression in Reverse Transcription Polymerase Chain

CC Reaction (RT-PCR) analysis. The methods are diagnostic techniques useful

CC for detecting and monitoring the progression of benign prostatic

CC hyperplasia and human prostate cancer (the most prevalent form of cancer

CC and a major cause of death in males) prior to the tumor undergoing

CC metastasis, therefore allowing the optimal method of treatment to be

CC determined before the condition becomes life threatening. The present

CC sequence represents a claimed marker gene fragment

XX Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

Query Match 13.2%; Score 275.4; DB 2; Length 757;

Best Local Similarity 96.2%; Pred. No. 9.3e-53;

Matches 282; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAACATTTGGATTTTTCCTTAACGCT 1745

DB 1 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAACATTTGGATTTTTCCTTAACGCT 60

QY 1746 AGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGATTTCCTC 1805

DB 61 AGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGATTTCCTC 120

QY 1806 CCTGTTATTAAGCAGCAAGCAAAATAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865

DB 121 CCTGTTATTAAGCAGCAAGCAAAATAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

QY 1866 TTCTTTTCCAAAGGAACCTCTAGAAGACCAAAATGCCCGAGTTAAGAACATCAAAAACCTAAC 1925

DB 181 TTCTTTTCCAAAGGAACCTCTAGAAGACCAAAATGCCCGAGTTAAGAACATCAAAAACCTAAC 240

QY 1926 CATCTGAAGAAACTTCCCAAGTGTAAAGACTCTGCCCTGCACGACAAACATATAA 1978

DB 241 CATCTGAAGAAACTTCCCAAGTGTAAAGACTCTGCCCTGCACGACAAACATATAA 293

RESULT 7

AAZ87503

ID AAZ87503 standard; cDNA; 757 BP.

AC AAZ87503;

XX AAZ87503;

XX AAZ87503;

DT 19-APR-2000 (first entry)

XX Prostate, breast and bladder cancers detecting biomarker UC Band #28.

XX Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;

XX benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;

XX diagnosis; ss.

XX Homo sapiens.

XX WO9964631-A1.

XX 16-DEC-1999.

XX 11-JUN-1999; 99WO-US013151.

XX 12-JUN-1998; 98US-00097199.

XX (UROC-) UROCOR INC.

XX An G, O'hara SM, Ralph D, Veltri RW;

XX WPI; 2000-116557/10.

XX Novel RNA biomarkers for diagnosis, prognosis and management of prostate,

XX breast and bladder cancer.

XX Claim 1; Page 165; 191pp; English.

XX The invention provides nucleic acid markers of prostate, breast and

XX bladder cancer. The markers are indicators of malignant transformation of

XX prostate, breast and bladder tissues and are diagnostic of the potential

XX for metastatic spread of malignant prostate tumours. The nucleic acid can

XX also be used as targets for therapeutic intervention in prostate cancer,

XX benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The

XX markers may be used to design specific probes and primers for the rapid

XX analysis of prostate, bladder or breast biopsy samples. The probes and

XX primers may also be used for in situ hybridization or in situ PCR

CC detection and diagnosis. They may also be used to identify and isolate

CC full length gene sequences from various DNA libraries. Antibodies against

CC the polypeptide products of the markers can be used to treat prostate

CC cancer, bladder cancer or breast cancer. The encoded proteins may be used

CC to detect antibodies. The proteins and antibodies can be used in

CC immunodetection methods for detecting or quantifying the cancers, and for

CC clinical diagnosis of these cancers. The antibodies may also be used for

CC radioimaging to quantify and localize the encoded proteins

XX Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

Query Match 13.2%; Score 275.4; DB 3; Length 757;

Best Local Similarity 96.2%; Pred. No. 9.3e-53;

Matches 282; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAACATTTGGATTTTTCCTTAACGCT 1745

DB 1 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAACATTTGGATTTTTCCTTAACGCT 60

QY 1746 AGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTAAGAGAGATTTCCTC 1805

DB 61 AGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTAAGAGAGATTTCCTC 120

QY 1806 CCTGTTATTAAGCAGCAAGCAAAATAGCCATTTCACTCTCAAACTTCACTAATGATCACA 1865

DB 121 CCTGTTATTAAGCAGCAAGCAAAATAGCCATTTCACTCTCAAACTTCACTAATGATCACA 180

QY 1866 TTCTTTTCCAAAGGAACCTCTAGAAGACCAAAATGCCCGAGTTAAGAACATCAAAAACCTAAC 1925

DB 181 TTCTTTTCCAAAGGAACCTCTAGAAGACCAAAATGCCCGAGTTAAGAACATCAAAAACCTAAC 240

QY 1926 CATCTGAAGAAACTTCCCAAGTGTAAAGACTCTGCCCTGCACGACAAACATATAA 1978

DB 241 CATCTGAAGAAACTTCCCAAGTGTAAAGACTCTGCCCTGCACGACAAACATATAA 293

RESULT 8

AAZ87522

ID AAZ87522 standard; cDNA; 757 BP.

AC AAZ87522;

XX AAZ87522;

DT 29-AUG-2001 (first entry)

XX Biomarker UC band 28 #1, used in diagnosis and prognosis of cancer.

XX Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;

XX benign prostatic hyperplasia; BPH; therapeutic; human; ss.

XX Homo sapiens.

XX US6218529-B1.

XX 17-APR-2001.

XX 12-JUN-1998; 98US-00097199.

XX 31-JUL-1995; 95US-0001655P.

XX 11-JAN-1996; 96US-0013611P.

XX 31-JUL-1996; 96US-00692787.

XX (UROC-) UROCOR INC.

XX An G, O'hara SM, Ralph D, Veltri R;

XX WPI; 2001-289849/30.

XX New nucleic acids as biomarkers and targets useful for detecting,

XX diagnosing, prognosing, and in developing treatments for prostate, breast

XX and bladder cancer.

XX Claim 2; Col 79; 78pp; English.

XX

[illegible]

PT biodiversity.
XX
PS Claim 1; SEQ ID NO 15140; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4051 BP; 1562 A; 824 C; 819 G; 846 T; 0 U; 0 Other;
Query Match 5.5%; Score 115.6; DB 5; Length 4051;
Best Local Similarity 89.9%; Pred. No. 4.3e-16;
Matches 124; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1951 AGACTCTGCTGCACGACACATATAAAGAGAGAGAGATCAATAGACACATATAA 2010
DB 1807 ATACTCTTTGAGAGAGCAACTCCAAAGAAAAGAGAGAGATCAATAGACACATATAA 1866
QY 2011 AAATGATAAAGGGGATATCACACCGATCCACAGAAATACAACTTACCATCAGAGAATA 2070
DB 1867 AAATGATAAAGGGGATATCACACCGATCCACAGAAATACAACTTACCATCAGAGAATA 1926
QY 2071 CTACAAACCTCTACGC 2088
DB 1927 CTACAAACCTCTACGC 1944
RESULT 11
ABT07086/c
ID ABT07086 standard; cDNA; 430 BP.
XX
AC ABT07086;
XX
DT 07-NOV-2002 (first entry)
XX
DE Human ovarian cancer associated coding sequence SEQ ID NO: 248.
XX
KW Human; ovarian cancer; cancer; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002076715-A1.
XX
PD 20-JUN-2002.
XX
PF 06-JUN-2001; 2001US-00876889.
XX
PR 23-SEP-1998; 98US-00159320.
XX
PR 08-FEB-1999; 99US-00246429.
XX
PR 16-SEP-1999; 99US-00397787.
XX
PA (BENS/) BENSON D R.
PA (LODE/) LODES M J.
PA (MITC/) MITCHAM J L.
PA (KING/) KING G E.

XX Benson DR, Lodes MJ, Mitcham JL, King GE;
XX WPI; 2002-598720/64.
XX
XX Composition for detecting and treating ovarian cancer, comprises a specific polypeptide, polynucleotide, T cell population, or antigen presenting cell.
XX
XX Example 1; Page 94; 189pp; English.
XX
XX The present invention relates to a method of detecting the presence of ovarian cancer in a patient, involving detecting ovarian cancer associated polynucleotides. The method is not only used to detect the presence of cancer, preferably ovarian cancer in a patient, but also is used to stimulate and/or expand T cells specific for an ovarian tumour protein. The sequences can be used in vaccines used to treat cancer. The present sequence is an ovarian cancer associated coding sequence
XX
SQ Sequence 430 BP; 70 A; 74 C; 85 G; 196 T; 0 U; 5 Other;
Query Match 5.5%; Score 115.2; DB 6; Length 430;
Best Local Similarity 97.5%; Pred. No. 2.7e-16;
Matches 117; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1969 AACCATATAAAGAGAGAGAGATCAATAGACACATATAAAGGGGATAT 2028
DB 235 AATAAAGAAAAGAGAGAGAGATCAATAGACACATATAAAGGGGATAT 236
QY 2029 CACCACCGATCCACAGAAATACAACTTACCATCAGAGAATACTACAAACCTCTACGC 2088
DB 235 CACCACCGATCCACAGAAATACAACTTACCATCAGAGAATACTACAAACCTCTACGC 176
RESULT 12
ABX72964/c
ID ABX72964 standard; cDNA; 430 BP.
XX
AC ABX72964;
XX
DT 14-MAR-2003 (first entry)
XX
DE Human ovarian carcinoma antigen partial cDNA sequence #230.
XX
KW Human; cancer detection; ovarian carcinoma antigen; ovarian cancer; tumour antigen; tumour; OV2; OV3; OV6; OV9; OV10; OV12; OV14; OV17; OV18; OV23; OV24; OV27; OV41; OV54; OV57; gene; ss.
XX
OS Homo sapiens.
XX
PN US6468758-B1.
XX
PD 22-OCT-2002.
XX
PF 16-SEP-1999; 99US-00397787.
XX
PR 23-SEP-1998; 98US-00159320.
XX
PR 08-FEB-1999; 99US-00246429.
XX
PA (CORI-) CORIXA CORP.
XX
PI Benson DR, Lodes MJ, Mitcham JL, King GE;
XX
XX WPI; 2003-147101/14.
XX
XX Determining presence or absence of cancer in patient by contacting patient sample with oligonucleotide that hybridizes to polynucleotide encoding ovarian carcinoma antigen, and detecting amount of hybridization in sample.
XX
XX Example 1; Col 181-182; 152pp; English.
XX
XX The present invention relates to compositions and methods for determining

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 06:16:18 ; Search time 99.5368 Seconds
(without alignments)
16226.276 Million cell updates/sec

Title: US-09-974-546C-83

Perfect score: 3589

Sequence: 1 gacctaataatcgaggt.....tactacaacacctctacgc 2088

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09974546/runat.07092005.174357.19482/app_query.fasta_1.4942
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SURFIX=n2p.rag -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODES=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CGN.1.1.232 @runat.07092005.174357.19482 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	694	19.3	135	3 AAY59296	Prostate
2	694	19.3	135	3 AAY59295	Prostate
3	694	19.3	135	4 AAU02174	Biomarker
4	694	19.3	135	4 AAU02175	Biomarker
5	206.5	5.8	717	4 ABG06939	Novel hum
6	206.5	5.8	1746	4 ABG05644	Novel hum
7	206.5	5.8	1824	4 ABG09152	Novel hum
8	204	5.7	724	4 ABG15384	Novel hum
9	204	5.7	724	4 ABG17531	Novel hum
10	204	5.7	1257	4 ABG17335	Novel hum

11	204	5.7	1921	4 ABG06576	Novel hum
12	203	5.7	625	4 ABG11195	Novel hum
13	203	5.7	626	4 ABG17324	Novel hum
14	202	5.6	590	4 ABG06851	Novel hum
15	202	5.6	857	4 ABG04403	Novel hum
16	202	5.6	954	4 ABG02322	Novel hum
17	202	5.6	954	4 ABG08451	Novel hum
18	202	5.6	956	4 ABG14785	Novel hum
19	202	5.6	956	4 ABG14541	Novel hum
20	202	5.6	956	5 ABG77047	Prostate
21	201.5	5.6	794	4 ABG17634	Novel hum
22	201.5	5.6	1557	4 ABG02512	Novel hum
23	201.5	5.6	1709	4 ABG03979	Novel hum
24	201.5	5.6	1709	4 ABG06055	Novel hum
25	201.5	5.6	1709	4 ABG19873	Novel hum
26	201.5	5.6	1709	4 ABG03052	Novel hum
27	201.5	5.6	1709	4 ABG10797	Novel hum
28	201.5	5.6	1709	4 ABG06602	Novel hum
29	201.5	5.6	1709	4 ABG20779	Novel hum
30	201.5	5.6	1709	4 ABG08635	Novel hum
31	201.5	5.6	1709	4 ABG02341	Novel hum
32	201.5	5.6	1709	4 ABG09638	Novel hum
33	201.5	5.6	1809	4 ABG01006	Novel hum
34	201	5.6	1129	8 ADP29913	Human sec
35	200	5.6	302	4 ABG14081	Novel hum
36	200	5.6	520	4 ABG03512	Novel hum
37	200	5.6	985	4 ABG06938	Novel hum
38	200	5.6	1065	4 ABG02332	Novel hum
39	200	5.6	1078	4 ABG10803	Novel hum
40	200	5.6	1078	4 ABG06060	Novel hum
41	200	5.6	1078	4 ABG06606	Novel hum
42	200	5.6	1078	4 ABG09642	Novel hum
43	200	5.6	1085	4 ABG14957	Novel hum
44	200	5.6	1196	4 ABG14610	Novel hum
45	200	5.6	1245	7 ADE09432	Novel pro

ALIGNMENTS

RESULT 1

AAV59296

ID AAY59296 standard; peptide; 135 AA.

XX

AC AAY59296;

XX

DT 19-APR-2000 (first entry)

XX

DE Prostate disease marker UC Band #28 amino acid sequence.

XX

KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer; benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.

XX

OS Homo sapiens.

XX

PN WO9964631-A1.

XX

PD 16-DEC-1999.

XX

PF 11-JUN-1999; 99WO-US013151.

XX

PR 12-JUN-1998; 98US-00097199.

XX

PA (UROC-) UROCOR INC.

XX

PI An G, O'hara SM, Ralph D, Veltri RW;

XX

DR WPI; 2000-116557/10.

XX

PT N-PSDB; AAZ87584.

XX

PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate, breast and bladder cancer.

XX

PS Example 5; Page 184-186; 191pp; English.

XX The invention provides nucleic acid markers of prostate, breast and
 CC bladder cancer. The markers are indicators of malignant transformation of
 CC prostate, breast and bladder tissues and are diagnostic of the potential
 CC for metastatic spread of malignant prostate tumours. The nucleic acid can
 CC also be used as targets for therapeutic intervention in prostate cancer,
 CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
 CC markers may be used to design specific probes and primers, for the rapid
 CC analysis of prostate, bladder or breast biopsy samples. The probes and
 CC primers may also be used for in situ hybridization or in situ PCR
 CC detection and diagnosis. They may also be used to identify and isolate
 CC full length gene sequences from various DNA libraries. Antibodies against
 CC the polypeptide products of the markers can be used to treat prostate
 CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
 CC to detect antibodies. The proteins and antibodies can be used in
 CC immunodetection methods for detecting or quantifying the cancers, and for
 CC clinical diagnosis of these cancers. The antibodies may also be used for
 CC radioimaging to quantify and localize the encoded proteins
 XX Sequence 135 AA;

Alignment Scores:
 Pred. No.: 1,14e-75 Length: 135
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 19.34% Indels: 0
 DB: 3 Gaps: 0

US-09-974-546C-83 (1-2088) x AAY59296 (1-135)

QY 99 ATGAGGCGCTTCTTAAGGAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
 Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisLeu 20
 QY 159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCGGCATTTAGCT 218
 Db 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 QY 219 CCAGAACTGTCTTCTTACCATTCTGCTACAGGTTATTCGAAAAAGAAAAAGTAAA 278
 Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGlnLysValLys 60
 QY 279 AGAAGTCAAAAGCGACAGATTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
 Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 QY 339 CTCACACCTTCACACACACATTCGACATGAAAGGTTCTCCTCAATGAATGTTCTCATTA 398
 Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 QY 399 TCTTTCAGAACCATATTATTCACATTGACTTTGCGAGTTAACTCAGACCTTAGGTCTGGA 458
 Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 QY 459 TCTCTCTTCTTCTACTTATCCAAAACATATACATCCACAGATCATTA 503
 Db 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

RESULT 2

AAY59295
 ID AAY59295 standard; peptide; 135 AA.

AC AAY59295;

XX AAY59295;

DT 19-APR-2000 (first entry)

XX Prostate disease marker UC Band #28 amino acid sequence.
 DE Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
 KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
 XX Homo sapiens.

XX WO9964631-A1.
 XX 16-DEC-1999.
 PD
 XX 11-JUN-1999; 99WO-US013151.
 PF
 XX 12-JUN-1998; 98US-00097199.
 PR
 XX (UROC-) UROCOR INC.
 PA
 XX An G, O'hara SM, Ralph D, Veltri RW;
 PI WPI: 2000-116557/10.
 XX N-PSDB; AA287583.
 DR
 XX Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
 PT breast and bladder cancer.
 PT
 XX Example 5; Page 182-183; 191pp; English.
 PS
 XX The invention provides nucleic acid markers of prostate, breast and
 CC bladder cancer. The markers are indicators of malignant transformation of
 CC prostate, breast and bladder tissues and are diagnostic of the potential
 CC for metastatic spread of malignant prostate tumours. The nucleic acid can
 CC also be used as targets for therapeutic intervention in prostate cancer,
 CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
 CC markers may be used to design specific probes and primers, for the rapid
 CC analysis of prostate, bladder or breast biopsy samples. The probes and
 CC primers may also be used for in situ hybridization or in situ PCR
 CC detection and diagnosis. They may also be used to identify and isolate
 CC full length gene sequences from various DNA libraries. Antibodies against
 CC the polypeptide products of the markers can be used to treat prostate
 CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
 CC to detect antibodies. The proteins and antibodies can be used in
 CC immunodetection methods for detecting or quantifying the cancers, and for
 CC clinical diagnosis of these cancers. The antibodies may also be used for
 CC radioimaging to quantify and localize the encoded proteins
 XX Sequence 135 AA;

Alignment Scores:
 Pred. No.: 1,14e-75 Length: 135
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 19.34% Indels: 0
 DB: 3 Gaps: 0

US-09-974-546C-83 (1-2088) x AAY59295 (1-135)

QY 99 ATGAGGCGCTTCTTAAGGAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
 Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisLeu 20
 QY 159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCGGCATTTAGCT 218
 Db 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 QY 219 CCAGAACTGTCTTCTTACCATTCTGCTACAGGTTATTCGAAAAAGAAAAAGTAAA 278
 Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGlnLysValLys 60
 QY 279 AGAAGTCAAAAGCGACAGATTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
 Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 QY 339 CTCACACCTTCACACACACATTCGACATGAAAGGTTCTCCTCAATGAATGTTCTCATTA 398
 Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 QY 399 TCTTTCAGAACCATATTATTCACATTGACTTTGCGAGTTAACTCAGACCTTAGGTCTGGA 458
 Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 QY 459 TCTCTCTTCTTCTACTTATCCAAAACATATACATCCACAGATCATTA 503
 Db 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 9, 2005, 11:23:58 ; Search time 21.589 Seconds
(without alignments)
14439.486 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 3589
Sequence: 1 gacctaataatagaggt.....tactacaacaccttaagc 2088

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0974546/runat.07092005.174359.19522/app.query.fasta_1.4942
-DB=Issued Patents AA -OPMT=faetan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLALIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0974546 @CGN 1.1.42 @runat.07092005.174359.19522 -NCPV=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfilees1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	771	21.5	157	4	US-09-949-016-9182 Sequence 9182, Ap
2	694	19.3	135	3	US-09-097-199-84 Sequence 84, Appl
3	694	19.3	135	3	US-09-097-199-86 Sequence 86, Appl
4	117.5	3.3	57	4	US-09-513-999C-7461 Sequence 7461, Ap
5	103	2.9	21	3	US-09-097-199-56 Sequence 56, Appl
6	100	2.8	3169	2	US-08-477-451-6 Sequence 6, Appl
7	91.5	2.5	944	3	US-09-134-001C-4352 Sequence 12, Appl
8	90.5	2.5	1720	2	US-08-477-451-12 Sequence 12, Appl
9	89	2.5	2184	4	US-09-417-485D-6 Sequence 6, Appl
10	88	2.5	305	4	US-09-328-352-6445 Sequence 6445, Ap
11	87	2.4	2210	4	US-09-309-572-7 Sequence 7, Appl
12	87	2.4	2210	4	US-09-718-096-7 Sequence 7, Appl

13	85	2.4	192	4	US-09-270-767-34244 Sequence 34244, A
14	85	2.4	192	4	US-09-270-767-43461 Sequence 43461, A
c 15	84.5	2.4	944	3	US-09-134-001C-4352 Sequence 4352, Ap
16	83	2.3	698	4	US-09-538-092-151 Sequence 151, App
c 17	82	2.3	318	4	US-09-393-634-47 Sequence 47, Appl
c 18	82	2.3	318	4	US-09-949-016-9253 Sequence 9253, Ap
19	81.5	2.3	944	4	US-09-107-532A-4864 Sequence 4864, Ap
20	81	2.3	128	6	5514582-34 Patent No. 5514582
21	81	2.3	128	6	5514582-34 Patent No. 5514582
22	80	2.2	979	4	US-09-538-092-482 Sequence 482, App
23	79.5	2.2	373	4	US-09-270-767-62380 Sequence 62380, A
24	79.5	2.2	671	4	US-09-270-767-45759 Sequence 46759, A
25	79.5	2.2	1151	4	US-09-328-352-5876 Sequence 5876, Ap
26	78.5	2.2	336	3	US-08-804-166-8 Sequence 8, Appl
27	78.5	2.2	336	3	US-08-910-991-8 Sequence 8, Appl
28	78.5	2.2	336	4	US-09-756-186-8 Sequence 8, Appl
29	78.5	2.2	998	4	US-09-949-016-8326 Sequence 8326, Ap
30	78	2.2	180	4	US-09-134-000C-6251 Sequence 6251, Ap
31	77.5	2.2	251	4	US-09-248-796A-24724 Sequence 24724, A
32	77	2.1	441	4	US-09-134-000C-6563 Sequence 6563, Ap
33	77	2.1	515	4	US-09-635-872A-6 Sequence 6, Appl
34	77	2.1	515	4	US-09-636-077A-6 Sequence 6, Appl
35	77	2.1	515	4	US-09-636-060C-6 Sequence 6, Appl
36	77	2.1	515	4	US-09-986-552-6 Sequence 6, Appl
37	77	2.1	515	4	US-09-636-596C-6 Sequence 6, Appl
38	77	2.1	515	4	US-10-023-894-18 Sequence 18, Appl
39	77	2.1	515	4	US-10-306-686-6 Sequence 6, Appl
40	77	2.1	666	1	US-08-318-831-3 Sequence 3, Appl
41	77	2.1	814	1	US-08-318-831-2 Sequence 2, Appl
42	76.5	2.1	339	4	US-09-248-796A-25501 Sequence 25501, A
43	76.5	2.1	369	4	US-09-248-796A-18582 Sequence 18582, A
c 44	76	2.1	125	4	US-09-270-767-59144 Sequence 59144, A
45	76	2.1	166	4	US-09-252-991A-18131 Sequence 18131, A

ALIGNMENTS

RESULT 1

US-09-949-016-9182
; Sequence 9182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9182
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9182

Alignment Scores:
Pred. No.: 6.22e-85 Length: 157
Score: 771.00 Matches: 155
Percent Similarity: 97.48% Conservative: 0
Best local Similarity: 97.48% Mismatches: 2
Query Match: 21.48% Indels: 2
DB: 4 Gaps: 0

US-09-974-546C-83 (1-2088) x US-09-949-016-9182 (1-157)

QY 31 TGTATAAATTAATTAACAAATTAATTTCTTATTGCTACAGACTACAAATTTACAGT 90

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 9, 2005, 12:21:48 ; Search time 117.944 Seconds
(without alignments)
13965.459 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 3589
Sequence: 1 Gaccttaatatatcgaggt.....tactacaacaccttaogc 2088

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1777461 seqs, 394431504 residues
Total number of hits satisfying chosen parameters: 3554922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09974546/runat_07092005_174400_19612/app_query.fasta_1.4942
-DB=Published Applications AA -QFMT=fastan -SUPPLX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09974546@cgn 1 229 @runat 07092005_174400_19612
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pcp.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pcp.*
20: /cgn2_6/ptodata/2/pubpaa/US11B_PUBCOMB.pcp.*
21: /cgn2_6/ptodata/2/pubpaa/US11C_PUBCOMB.pcp.*
22: /cgn2_6/ptodata/2/pubpaa/US11D_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	694	19.3	135	10	US-09-974-546-84
2	694	19.3	135	10	US-09-974-546-86
3	202	5.6	956	14	US-10-000-256A-153
4	197	5.5	197	13	US-10-001-835-189
5	197	5.5	1031	14	US-10-082-828A-257
6	197	5.5	1031	20	US-11-057-807-257
7	197	5.5	1275	14	US-10-025-201-3
8	197	5.5	1275	17	US-10-800-322-34
9	195	5.4	1397	17	US-10-473-451-6
10	192	5.3	1010	14	US-10-050-882-80
11	192	5.3	1272	15	US-10-114-270-30
12	192	5.3	1274	16	US-10-415-615-3
13	191.5	5.3	1081	15	US-10-085-198-150
14	191	5.3	936	15	US-10-243-552-850
15	191	5.3	1018	15	US-10-085-198-156
16	189	5.3	843	15	US-10-001-885-115
17	189	5.3	843	20	US-11-057-447-115
18	183	5.1	990	15	US-10-085-198-154
19	174	4.8	190	14	US-10-074-475-250
20	172	4.8	4961	14	US-10-114-153-64
21	170	4.7	940	14	US-10-082-830-235
22	125	3.5	316	13	US-10-001-870-132
23	122	3.4	100	15	US-10-094-749-2789
24	118	3.3	112	11	US-09-864-408A-6702
25	111.5	3.1	127	15	US-10-094-749-3226
26	106	3.0	1017	15	US-10-085-198-152
27	103.5	2.9	371	15	US-10-369-493-20398
28	103.5	2.9	381	15	US-10-369-493-1334
29	103	2.9	21	10	US-09-974-546-56
30	93.5	2.6	6761	17	US-10-732-923-15035
31	91.5	2.5	461	16	US-10-748-112-21
32	91	2.5	35	14	US-10-105-232-510
33	91	2.5	35	14	US-10-189-437-497
34	91	2.5	39	14	US-10-105-232-508
35	91	2.5	39	14	US-10-189-437-495
36	91	2.5	40	14	US-10-105-232-518
37	91	2.5	40	14	US-10-189-437-505
38	91	2.5	44	14	US-10-105-232-514
39	91	2.5	44	14	US-10-105-232-517
40	91	2.5	44	14	US-10-189-437-501
41	91	2.5	44	14	US-10-189-437-504
42	91	2.5	48	14	US-10-105-232-513
43	91	2.5	48	14	US-10-189-437-500
44	91	2.5	56	14	US-10-105-232-507
45	91	2.5	56	14	US-10-189-437-494

ALIGNMENTS

RESULT 1
US-09-974-546-84
; Sequence 84, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert

TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE

NUMBER OF SEQUENCES: 87

CORRESPONDENCE ADDRESS:

ADDRESS: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/974,546
/ FILING DATE: 10-Oct-2001
/ CLASSIFICATION: Unknown
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/097,199
/ FILING DATE: 1998-06-12
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 84:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 135 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-974-546-84

Alignment Scores:
Pred. No.: 7,85e-71 Length: 135
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.34% Indels: 0
DB: 10 Gaps: 0

US-09-974-546C-83 (1-2088) x US-09-974-546-84 (1-135)

QY 99 ATGAGGGCTTCTTAAGGAACAGAAATATGAGGATATGCAATATATTCACATTTTA 158
Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20
QY 159 CAGATCAGAAAATTGAGGCACAGATTAACTCTCCCAAGGCTACAGGCAATTCAGCT 218
Db 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
QY 219 CCAGAACTGTCTTACCATTCTGCTACAGGTTTCGAAAAAGAAAAAGCTAAA 278
Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
QY 279 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTACACCATGCAATT 338
Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
QY 339 CTCACACCTTCGAGACACACTTGACCATGAAAGTTCTCTCAATGAAATGTTCCTCATTA 398
Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerMetLysCysSerLeu 100
QY 399 TCTTCAGAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCCTAGGCTCGAA 458
Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
QY 459 TCTGTCTTCTTACTTATCCAAAACATATACATCCACAGATCATTA 503
Db 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

RESULT 2

US-09-974-546-86
/ Sequence 86, Application US/09974546
/ Publication No. US20030050470A1
/ GENERAL INFORMATION:
/ APPLICANT: An, Gang
/ O'Hara, S. Mark
/ Ralph, David

/ Veltri, Robert
/ TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS
/ PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
/ NUMBER OF SEQUENCES: 87
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Arnold, White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/974,546
/ FILING DATE: 10-Oct-2001
/ CLASSIFICATION: Unknown
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/097,199
/ FILING DATE: 1998-06-12
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 86:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 135 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

Alignment Scores:
Pred. No.: 7,85e-71 Length: 135
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.34% Indels: 0
DB: 10 Gaps: 0

US-09-974-546C-83 (1-2088) x US-09-974-546-86 (1-135)

QY 99 ATGAGGGCTTCTTAAGGAACAGAAATATGAGGATATGCAATATATTCACATTTTA 158
Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20
QY 159 CAGATCAGAAAATTGAGGCACAGATTAACTCTCCCAAGGCTACAGGCAATTCAGCT 218
Db 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
QY 219 CCAGAACTGTCTTACCATTCTGCTACAGGTTTCGAAAAAGAAAAAGCTAAA 278
Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
QY 279 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTACACCATGCAATT 338
Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
QY 339 CTCACACCTTCGAGACACACTTGACCATGAAAGTTCTCTCAATGAAATGTTCCTCATTA 398
Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerMetLysCysSerLeu 100
QY 399 TCTTCAGAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCCTAGGCTCGAA 458
Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 9, 2005, 09:43:33 ; Search time 25.2251 Seconds
(without alignments)
15928.642 Million cell updates/sec

Title: US-09-974-546C-83

Perfect score: 3589

Sequence: 1 gacccataatcatcgaggt.....tactacaaacacctctaacg 2088

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09974546/runat.07092005.174358.19508/app_query.fasta_1.4942
-DB=pir.79 -OPMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOCPCL=0 -LOCPXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CGN 1.1 79 @runat.07092005.174358.19508 -NCPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR.79:*

1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	197	5.5	1275	2 B28096	line-1 protein ORF
2	197	5.5	1275	2 I38588	reverse transcript
3	197	5.5	1275	2 S65824	reverse transcript
4	195	5.4	1259	4 GNHUL1	retrovirus-related
5	187	5.2	1280	2 B34087	hypothetical protein
6	171.5	4.8	138	2 I59767	lamin-like protein
7	156	4.3	335	4 A26718	retrovirus-related
8	126.5	3.5	1260	4 GNRLR1	retrovirus-related
9	125	3.5	190	2 B25556	hypothetical protein
10	124	3.5	1281	1 GNMSLL	retrovirus-related
11	118	3.3	685	2 S16783	probable RNA-direc
12	115	3.2	500	2 S16788	probable reverse t
13	114	3.2	556	2 S21347	hypothetical prote
14	103.5	2.9	381	2 F71196	probable hexosyltr

ALIGNMENTS

RESULT 1

B28096

line-1 protein ORF2 - human

C:Species: Homo sapiens (man)

C>Date: 03-Nov-1988 #sequence_revision 03-Nov-1988 #text_change 09-Jul-2004

C:Accession: B28096

R:Skowronski, J.; Fanning, T.G.; Singer, M.F.

Mol. Cell. Biol. 8, 1395-1397, 1988

A:Title: Unit-length line-1 transcripts in human teratocarcinoma cells.

A:Reference number: A28096; MUID:88246405; PMID:2454389

A:Accession: B28096

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1275 <SKO>

A:Cross-references: UNIPROT:Q15604; UNIPROT:QJUN80; UNIPROT:Q12881; UNIPROT:O00363; UNIPROT:O00375

PROT:Q9Y5K0; UNIPROT:O00366; UNIPROT:Q8TE30; UNIPROT:O00375

C:Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	1,058-12	Length:	1275
Score:	197.00	Matches:	37
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.49%	Indels:	0
DB:	2	Gaps:	0

US-09-974-546C-83 (1-2088) x B28096 (1-1275)

QY 1976 AAAAAAGAGAGAAGAAATCAATAGACACATAAAAAATGATAAAGGGGATATCACCACC 2035

383 LysLysArgGluLysAsnGlnLeuThrLeuLysAsnAspLysGlyAspLeuThr 402

2036 GATCCACAGAAATCAAACTACCATCAGAGAACTACTACAAACACCTCTAC 2086

403 AspProThrGluLeuGlnThrThrLeuArgGluTyTyLysHisLeuTy 419

RESULT 2

I38588

reverse transcriptase homolog - human retrotransposon L1

N/Alternate names: ORF2 protein

C/Species: Homo sapiens (man)

C/Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004

C/Accession: I38588

R/Holmes, S.B.; Dombroski, B.A.; Krebs, C.M.; Boehm, C.D.; Kazazian, H.H.

Nature Genet. 7, 143-148, 1994

A/Title: A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q p

A/Reference number: I38587; MUID:95004577; PMID:7920631

A/Accession: I38588

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1275 <RES>

A/Cross-references: UNIPROT:Q12881; EMBL:U09116; NID:G483914; PIDN:ARB60345.1; PID:G4839

C/Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	1.05e-12	Length:	1275
Score:	197.00	Matches:	37
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.49%	Indels:	0
DB:	2	Gaps:	0

US-09-974-546C-83 (1-2088) x I38588 (1-1275)

QY 1976 AAAAAAGAGAGAAGTAATCAATAGACACATAAAAAATGATAAGGGGATATCACCACC 2035

DB 383 LysLysArgGluLysAsnGlnIleAspThrIleLysAsnAspGlyAspIleThrThr 402

QY 2036 GATCCACAGAATAACAACTACCATCAGAGAATACTACAAACACCTCTAC 2086

DB 403 AspProThrGluIleGlnThrThrIleArgGluTyTyTyTyLysHisLeuTy 419

RESULT 3

S65824

reverse transcriptase homolog - human transposon L1.1

C/Species: Homo sapiens (man)

C/Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004

C/Accession: S65824

R/Dombroski, B.A.

submitted to the EMBL Data Library, January 1992

A/Description: Isolation of an active human transposable element.

A/Reference number: S65823

A/Accession: S65824

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1275 <DOM>

A/Cross-references: UNIPROT:Q15604; EMBL:M80340; NID:G339767; PIDN:AAA51622.1; PID:G3397

C/Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	1.05e-12	Length:	1275
Score:	197.00	Matches:	37
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.49%	Indels:	0
DB:	2	Gaps:	0

US-09-974-546C-83 (1-2088) x S65824 (1-1275)

QY 1976 AAAAAAGAGAGAAGTAATCAATAGACACATAAAAAATGATAAGGGGATATCACCACC 2035

DB 383 LysLysArgGluLysAsnGlnIleAspThrIleLysAsnAspGlyAspIleThrThr 402

QY 2036 GATCCACAGAATAACAACTACCATCAGAGAATACTACAAACACCTCTAC 2086

DB 403 AspProThrGluIleGlnThrThrIleArgGluTyTyTyTyLysHisLeuTy 419

RESULT 4

GNHUL1

retrovirus-related reverse transcriptase pseudogene - human

C/Species: Homo sapiens (man)

C/Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 09-Jul-2004

C/Date: 31-Mar-1988 #sequence_revision 04-Jan-1996 #text_change 09-Jul-2004

C/Accession: A25313

R/Hattori, M.; Kuhara, S.; Takenaka, O.; Sakaki, Y.

Nature 321, 625-628, 1986

A/Title: L1 family of repetitive DNA sequences in primates may be derived from a sequence

A/Reference number: A93381; MUID:86230917; PMID:2423883

A/Accession: A25313

A/Status: conceptual translation of pseudogene

A/Molecule type: DNA

A/Residues: 1-1259 <HAT>

A/Cross-references: UNIPROT:P08547

A/Note: this sequence was constructed from an alignment of published and unpublished seq

C/Keywords: reverse transcriptase; pseudogene

Alignment Scores:

Pred. No.:	1.74e-12	Length:	1259
Score:	195.00	Matches:	69
Percent Similarity:	47.52%	Conservative:	27
Best Local Similarity:	34.16%	Mismatches:	77
Query Match:	5.43%	Indels:	29
DB:	4	Gaps:	8

US-09-974-546C-83 (1-2088) x GNHUL1 (1-1259)

QY 1501 AAACCTCAGTAAGATATATATTAGTCTGTTGCAAAGTAGAAAAAGATTCTCATCTCAAC 1560

DB 240 LysLeuThrGlnAsnHisSerThrThrTrp---LysLeuAsnAsnLeuLeuLeuAsnAsp 258

QY 1561 CTTATGACGAGGAAGAG-----GGAAGGCTGTTTGAGAACCATTTACTT 1605

DB 259 TyrTrpValHisAsnGluMetLysAlaGluIleLysLysPheGluThrAsnGluAsn 278

QY 1606 AGCAGAACCATATTTTAGACACTTCCTGCATTACCTGCACAAACATAATGTTTG--- 1662

DB 279 LysAspThrThrTyrglnAsnLeuTrpAsp-----ThrAlaLysAlaValCysArgGly 296

QY 1663 CAACTTGTGTGATCAACCTCCACACAGCACATTCAGAGATTAAATATTATTTTCATCAA 1722

DB 297 LysPheIleAlaLeuAsnAlaHisLysArgLysGlnGluArgSerLysIle-----Asp 314

QY 1723 ACATTGGATTTTTCCTTAACGCTAGAGATTGCTACAAATCTTCTGAAGGTTCTCAATGGC 1782

DB 315 ThrLeu---IleSerGlnLeuLysGluLeuGluLysGlnGluThrAsn-SerLysAl 333

QY 1783 TTCAGGCTAAGAAGAGATTCTTCCTGTTTATAAGCAGCAGCAACAATTTAGCCATTTCAC 1842

DB 333 aserArgArgGlnGluIleIle-----LysIleArgAlaGluLe 346

QY 1843 CTCAAACTTCTCATATGATCATCTTTTCCAAAAGGAACCTCTAGAGACCAAAATGCCCC 1902

DB 346 uLysGluIleGluThrGlnLysThrLeuGlnLysIleAsn----- 359

QY 1903 GAGTTAAGAACATCAAAATCAACATCTGAGAAATCTCCCAAGTGTAGACTCTGCCTG 1962

DB 360 -----GluSerArgSerTrpPheGluLysIleAsnLysIleAspArgProLeuAl 377

QY 1963 CACGACACACATAAAAAAGAGAGAATCAATAGACACATAAAAAATGATAAAGG 2022

DB 377 aArgLeuIleLysLysLysArgGluLysAsnGlnIleAspThrIleLysAsnAspArgGl 397

QY 2023 GGATATCACCCGATCCACAGAAATACAACTACCATCAGAGAAATACATAACACACT 2082

DB 397 yAspIleThrThrAspProThrGluIleGlnThrThrIleArgGluTyTyTyLysHisLe 417

QY 2083 CTAC 2086

DB 417 uTyr 418

RESULT 5

B34087

hypothetical protein (L1H 3' region) - human

C/Species: Homo sapiens (man)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 9, 2005, 06:29:13 ; Search time 124.307 Seconds
(without alignments)
17202.865 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 3589
Sequence: 1 gacctaataatagaggt.....tactacaaacactctacgc 2088

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DBV=xlh
-O=/cgn2 1/USPTO.spool/US09974546/runat.07092005 174358 19494/app query.fasta 1.4942
-DB=uniprot_03 -QPM=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09974546 @CGN 1 1 302 @runat.07092005 174358 19494 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : uniprot_03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	19.3	135	2 Q9GZY1	Q9gzy1 homo sapien
2	200	5.6	1275	2 Q00372	Q00372 homo sapien
3	198.5	5.5	1275	2 Q00375	Q00375 homo sapien
4	197	5.5	111	2 Q9UJ24	Q9uj24 homo sapien
5	197	5.5	1192	2 Q7KZ41	Q7kz41 homo sapien
6	197	5.5	1275	2 Q00360	Q00360 homo sapien
7	197	5.5	1275	2 Q00366	Q00366 homo sapien
8	197	5.5	1275	2 Q00368	Q00368 homo sapien
9	197	5.5	1275	2 Q00370	Q00370 homo sapien
10	197	5.5	1275	2 Q00378	Q00378 homo sapien
11	197	5.5	1275	2 Q8TE30	Q8te30 homo sapien
12	195	5.4	1259	1 LIN1 HUMAN	P08547 homo sapien
13	193.5	5.4	1275	2 Q00362	Q00362 homo sapien
14	190	5.3	1275	2 Q00363	Q00363 homo sapien
15	171.5	4.8	138	2 Q14755	Q14755 homo sapien
16	136	3.8	85	2 Q25910	Q25910 plasmodium

17	136	3.8	232	2 Q6ZNM9	Q6znm9 homo sapien
18	128	3.6	1356	2 Q6Q176	Q6qi76 rattus norv
19	126.5	3.5	1260	1 LIN1_NYCCO	P08548 nycticebus
20	125	3.5	190	2 Q63501	Q63501 rattus norv
21	124	3.5	1252	2 Q9QWY2	Q9qwy2 mus musculu
22	124	3.5	1281	2 O54850	O54850 mus musculu
23	124	3.5	1281	2 O88913	O88913 mus musculu
24	124	3.5	1281	2 O88915	O88915 mus musculu
25	124	3.5	1281	2 Q91288	Q91288 mus musculu
26	124	3.5	1281	2 Q79219	Q79219 mus musculu
27	124	3.5	1281	2 Q9QWY0	Q9qwy0 mus musculu
28	124	3.5	1281	2 Q9QWY0	Q9qwy0 mus musculu
29	124	3.5	1300	1 POL2_MOUSE	P11369 mus musculu
30	124	3.5	1300	2 O88906	O88906 mus musculu
31	123	3.4	1281	2 O88914	O88914 mus musculu
32	123	3.4	1281	2 Q9QWY3	Q9qwy3 mus musculu
33	121.5	3.4	1575	2 Q6TXJ1	Q6txj1 rattus norv
34	121	3.4	1275	2 O62658	O62658 canis famil
35	120	3.3	1281	2 Q91289	Q91289 mus musculu
36	119	3.3	484	2 Q7R7B9	Q7r7b9 plasmodium
37	119	3.3	931	2 Q6Q135	Q6qi35 rattus norv
38	119	3.3	1102	2 Q6TUD6	Q6tud6 rattus norv
39	119	3.3	1300	2 P97692	P97692 rattus norv
40	119	3.3	1366	2 Q6Q158	Q6qi58 rattus norv
41	119	3.3	2321	2 Q6TXE2	Q6txe2 rattus norv
42	118	3.3	685	2 Q63288	Q63288 rattus norv
43	118	3.3	1217	2 Q6TUF6	Q6tuf6 rattus norv
44	118	3.3	1219	2 Q9QYM3	Q9qym3 mus musculu
45	118	3.3	1300	2 Q61785	Q61785 mus musculu

ALIGNMENTS

RESULT 1
Q9GZY1 PRELIMINARY; PRT; 135 AA.
AC Q9GZY1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE UC28 protein.
GN Name=UC28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate cancer;
RX MEDLINE=21028101; PubMed=11156405;
RA An G., Ng A.Y., Meka C.S.R., Luo G., Bright S.P., Cazares L.,
RA Wright G.L. Jr., Veltri R.W.;
RT "Cloning and characterization of UROC28, a novel gene overexpressed in
RT prostate, breast, and bladder cancers.";
RL Cancer Res. 60:7014-7020(2000).
DR EMBL; AF189270; AAG17118.1; -.
DR EMBL; AF189269; AAG17117.1; -.
DR Genew; HGNC:21079; PBOV1.
SQ SEQUENCE 135 AA; 15722 MW; 2B7DB8B1983705D0 CRC64;

Alignment Scores:
Pred. No.: 2.48e-64 Length: 135
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.34% Indels: 0
DB: 2 Gaps: 0

US-09-974-546C-83 (1-2088) x Q9GZY1 (1-135)

QY 99 ATGAGGCGCTCTTAAGGACCCAGAAATATGAGGATATGCACATATTATTACATTTTA 158
|||||
DB 1 MetArgAlaPheLeuArgAsnGlnIstYrGluAspMetHisIleIleLeu 20
|||||

```
QY 159 CAGATCAGAAATGAGGCACAGATTAACTTCCCAAGGCTACCAAGCATCTTAGCT 218
Db 21 GlnileArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyileuAla 40
QY 219 CAGAAACTGTGCTCTTACATCTCTCTCAAGTATTCGAAAAAAGAAAGTAA 278
Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
QY 279 AGAAGTCAAAGGCACAGATTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
QY 339 CTCACACCCCTTGACACACATGACCATGAAAGGTTCTCAATGAAATGTCCTCATTA 398
Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
QY 399 TCTTCAGAGCCATATATTCATCATGATTTGCGAGTTAACTCAGACCCCTAGCTCGAA 458
Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
QY 459 TGCTGTCTTCTCTACTATCCAAACTATACATCCACAGATCATA 503
Db 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIle 135

RESULT 2
O00372 ID O00372 PRELIMINARY; PRT; 1275 AA.
AC O00372;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazanian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93570; AAC51273.1; -.
DR PIR; B28096; B28096.
DR PIR; S23650; S23650.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTase.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT_1; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149110 MW; 4711B3BC22F7674E CRC64;

Alignment Scores:
Pred. No.: 1.09e-11 Length: 1275
Score: 200.00 Matches: 64
Percent Similarity: 42.02% Conservativeness: 15
Best Local Similarity: 34.04% Mismatches: 47
Query Match: 5.57% Indels: 62
DB: 2 Gaps: 6

US-09-974-546c-83 (1-2088) x O00372 (1-1275)

QY 1523 GTCGTGTTCAAGTAGAAAAAGATTCTCATCTCACTCAACCTTATGAGCAGGAGGAA 1582
Db 294 ValCys-----ArgGlyLysPheThrAlaLeuAsnAlaTyrLysArgLysGlnGlu 310
QY 1583 GCGCTGTTGAGAACCATTTACTTAGCAGAACCAACATATTTTAGACACTTCCCTCGCATTA 1642
```

```
Db 311 -----ArgSerLysIleAspThrLeuThrSerGlnLeuLysGlu 323
QY 1643 CTGCACAAACAATATGTTTGGCAAACTGTTTGATCAACCTCAACAACGACACATTCAGG 1702
Db 324 LeuGluLysGln-----GluGlnThrHis--- 331
QY 1703 AGTTAAATATTTTTCATCAAAACATTTGGATTTTTCCTTAACGCTAGAGATTGCTACAAATC 1762
Db 331 ----- 331
QY 1763 TTCTGAAGGCTCTCAATGGCTTCAAGGCTAAGAAGAGATTTCCTCCCTGTTATAAGCAGCAA 1822
Db 332 -----SerLysAlaSerArgGlnGlnIleThr----- 341
QY 1823 GACAAATAGCCATTTCACTCTCAAACTTCACTAATGATCATTCTTTCCAAAGGAAC 1882
Db 342 ---LysIleArgAlaGluLeuLysGluLeuThrGlnLysThrLeuGlnLysIleAsn 360
QY 1883 TCTAGAGACCAAAATGCCCGAGTTTAAGAACATCAAACTAACCATCTCGAAGAACTTCC 1942
Db 361 -----GluSerArgSerTyrPhePheGluArgIleAsn 371
QY 1943 CAAGTGTAAGACTCTGCTGCACGACACATCAAAAGAGAGAAATCAAAATAGAC 2002
Db 372 LysIleAspArgProLeuAlaArgLeuIleLysLysLysArgGluLysAsnGlnIleAsp 391
QY 2003 ACAATAAAAAATGATAAAGGGGATATCACACCGATCCACAGAAATCAAACTACCATC 2062
Db 392 ThrIleLysAsnAspLysGlyAspIleThrThrAspProThrGluIleGlnThrThrIle 411
QY 2063 AGAGAATACTACAAACACCTCTAC 2086
Db 412 ArgGluTyrTyrLysHisLeuTyr 419

RESULT 3
O00375 ID O00375 PRELIMINARY; PRT; 1275 AA.
AC O00375;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative p150.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97285120; PubMed=9140393;
RA Sassaman D.M., Dombroski B.A., Moran J.V., Kimberland M.L., Naas T.P.,
RA DeBerardinis R.J., Gabriel A., Swergold G.D., Kazanian H.H. Jr.;
RT "Many human L1 elements are capable of retrotransposition.";
RL Nat. Genet. 16:37-43(1997).
DR EMBL; U93572; AAC51276.1; -.
DR PIR; B28096; B28096.
DR PIR; S23650; S23650.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000477; RVTase.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00078; RVT_1; 1.
KW RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1275 AA; 149054 MW; 456DD0F3DD7A17F5 CRC64;

Alignment Scores:
Pred. No.: 1.57e-11 Length: 1275
Score: 198.50 Matches: 76
Percent Similarity: 37.74% Conservativeness: 21
Best Local Similarity: 29.57% Mismatches: 75
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:24:09 ; Search time 6016.29 Seconds
(without alignments)
16816.746 Million cell updates/sec

Title: US-09-974-546C-83
Perfect score: 2088
Sequence: 1 gacctaataatcaggtggt.....tactacaacacctctacgc 2088

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sta.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2086	99.9	2103	9	AF189270 Homo sapi
2	2076	99.4	2087	6	AR146835 Sequence
3	1961.6	93.9	2512	9	AF189269 Homo sapi
4	1950	93.4	2505	6	AR146836 Sequence
5	1920.2	92.0	12927	9	HS171N11
6	1920.2	92.0	186658	2	AC036236 Homo sapi
7	1919.2	91.9	2496	9	BC069109 Homo sapi
8	275.4	13.2	757	6	AR146757 Sequence
9	129.2	6.2	230685	2	AC131318 Mus muscu
10	123.2	5.9	167998	2	AC026382 Mus muscu
11	122	5.8	110000	2	AC114782_1
12	122	5.8	163271	9	AC093897 Homo sapi
13	121	5.8	154941	9	AC117430 Homo sapi
14	121	5.8	155722	2	AC053487 Homo sapi
15	120.4	5.8	180388	9	HUMRETBLAS
16	120.4	5.8	203753	9	AC104343 Homo sapi
17	119.8	5.7	108175	9	AC068206 Homo sapi
18	119.8	5.7	159299	9	AC090257 Homo sapi
19	119.8	5.7	169571	9	AC020687 Homo sapi

20	119.8	5.7	209152	2	AC015558 Homo sapi
21	119.6	5.7	167961	2	AF002422 Homo sapi
c 22	119.4	5.7	95650	9	AC119737 Homo sapi
23	119.4	5.7	126838	9	AC125609 Homo sapi
24	119.4	5.7	169402	2	AC027715 Homo sapi
25	118.8	5.7	38528	9	HSAC001642 Human Chr
26	118.8	5.7	73589	9	AP001770 Homo sapi
27	118.8	5.7	109666	9	AP000838 Homo sapi
c 28	118.8	5.7	158485	9	AC099545 Homo sapi
29	118.8	5.7	160426	2	AC068372 Homo sapi
c 30	118.8	5.7	165870	9	AC025089 Homo sapi
31	118.8	5.7	174916	9	AC108056 Homo sapi
c 32	118.8	5.7	175560	9	AL499602 Human DNA
c 33	118.8	5.7	176802	2	AC012383 Homo sapi
34	118.8	5.7	184490	9	HSU82828 Homo sapien
c 35	118.8	5.7	186104	9	AP005718 Homo sapi
c 36	118.8	5.7	194149	9	AC100799 Homo sapi
37	118.8	5.7	196695	2	AC078936 Homo sapi
38	118.8	5.7	241802	2	AC147280 Pan trogl
39	117.8	5.6	161478	9	AC025538 Homo sapi
c 40	117.8	5.6	200923	9	AC091928 Homo sapi
c 41	117.2	5.6	40392	9	AC004610 Homo sapi
c 42	117.2	5.6	49743	9	AL450466 Human DNA
c 43	117.2	5.6	51407	9	AL357556 Human DNA
44	117.2	5.6	74625	9	AL596214 Human DNA
45	117.2	5.6	78607	9	AC145092 Pan trogl

ALIGNMENTS

RESULT 1
AF189270
LOCUS AF189270 2103 bp mRNA linear PRI 22-MAR-2001
DEFINITION Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION AF189270
VERSION AF189270.1 GI:10441603
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2103)
AUTHORS An.G., Ng.A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
Wright,G.L. Jr. and Veltre,R.W.
TITLE Cloning and characterization of UROC28, a novel gene overexpressed
in prostate, breast, and bladder cancers
JOURNAL Cancer Res. 60 (24), 7014-7020 (2000)
MEDLINE 21028101
PubMed 11156405
REFERENCE 2 (bases 1 to 2103)
AUTHORS An.G. and Veltre,R.W.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
Parkway, Oklahoma City, OK 73104, USA
FEATURES
source
1..2103
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q23-q24"
/tissue_type="prostate cancer"
1..2103
/gene="UC28"
99..506
/gene="UC28"
/codon_start=1
/product="UC28 protein"
/protein_id="AAG17118.1"
/db_xref="GI:10441604"
/translation="MRALRNQKYMENIHLQIRKLRHFLSNPRLPGILAPETY
LUPFCYKVRKKEKVRQKATEFIDYSIEQSHALLPLQTHLTKGSSMKCSLSS

EAIFLTLQLTQTTLGLECLLYLSKTHPQII"

misc_feature 1961..2103
/gene="UC28"
polyA_signal /note="alternative splicing"
2047..2052
/gene="UC28"

ORIGIN

Query Match		99.9%	Score 2086;	DB 9;	Length 2103;
Best Local Similarity		100.0%	Pred. No. 0;	Mismatches 0;	Indels 0; Gaps 0;
Matches 2087;		Conservative	0;	0;	0;
Qy	1	GACCTTAAATATATCAGAGTGGCTAAATTGATGTATATAATTTACAAAATTAATCTTCTTA	60		
Db	1	GACCTTAAATATATCAGAGTGGCTAAATTGATGTATATAATTTACAAAATTAATCTTCTTA	60		
Qy	61	TTGCTACAGAGCTACAATTCAAATTTACAGTAGGCCACCAATGAGGGCTCTTTAAGGAACC	120		
Db	61	TTGCTACAGAGCTACAATTCAAATTTACAGTAGGCCACCAATGAGGGCTCTTTAAGGAACC	120		
Qy	121	AGAAATATGAGGATATGCACAAATTTATTCACATTTTACAGATCAGAAAATTTGAGGCACA	180		
Db	121	AGAAATATGAGGATATGCACAAATTTATTCACATTTTACAGATCAGAAAATTTGAGGCACA	180		
Qy	181	GATTAAAGTAACTTCCCAAGGCTACCAAGGCATTTAGCTCCAGAACTGTGCTCTTACCAT	240		
Db	181	GATTAAAGTAACTTCCCAAGGCTACCAAGGCATTTAGCTCCAGAACTGTGCTCTTACCAT	240		
Qy	241	TCGTGTACAGGCTATTTCCGAAAAGAAAGTAAAGTAAAGTCAAAAGGCAACAGAGT	300		
Db	241	TCGTGTACAGGCTATTTCCGAAAAGAAAGTAAAGTAAAGTCAAAAGGCAACAGAGT	300		
Qy	301	TCATTGATTAATTCATAGAACAGTCAACCATGCAATTTCTCACACCCTTTCGACACACACT	360		
Db	301	TCATTGATTAATTCATAGAACAGTCAACCATGCAATTTCTCACACCCTTTCGACACACACT	360		
Qy	361	TGACCATGAAGGTTCCCTCAATGAATGTTCCTCATTAATCTTCAGAAAGCCATATTATCCA	420		
Db	361	TGACCATGAAGGTTCCCTCAATGAATGTTCCTCATTAATCTTCAGAAAGCCATATTATCCA	420		
Qy	421	CATTGCTTTGCGAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTCTCTACTATTATCCA	480		
Db	421	CATTGCTTTGCGAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTCTCTACTATTATCCA	480		
Qy	481	AAACTATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA	540		
Db	481	AAACTATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA	540		
Qy	541	ATAAAAATGGTTGAAAAGGCAATTCCTGCTACCAATGACTGTTTAAAGCCAGCCAAAGTAAC	600		
Db	541	ATAAAAATGGTTGAAAAGGCAATTCCTGCTACCAATGACTGTTTAAAGCCAGCCAAAGTAAC	600		
Qy	601	TGACCATTTCCAACTTCAATTTACTTTATGAAAAGAAATTTGATGATGATGAGGTTATTTC	660		
Db	601	TGACCATTTCCAACTTCAATTTACTTTATGAAAAGAAATTTGATGATGATGAGGTTATTTC	660		
Qy	661	AATTCCTAAAATACAAACCCATTTGATCTTTCTCAATCTTGAATCATAGATTAATATCT	720		
Db	661	AATTCCTAAAATACAAACCCATTTGATCTTTCTCAATCTTGAATCATAGATTAATATCT	720		
Qy	721	ATTATCTCAATTTAGTTTGTATTTATTCCTAGTGGGCCATTAATAAATACTACCATGTGTT	780		
Db	721	ATTATCTCAATTTAGTTTGTATTTATTCCTAGTGGGCCATTAATAAATACTACCATGTGTT	780		
Qy	781	TCGTCTCTCCATTTAGTCAATTAATTAATTAACGAGCAATTTAGTAGCCATGTCAGAT	840		
Db	781	TCGTCTCTCCATTTAGTCAATTAATTAATTAACGAGCAATTTAGTAGCCATGTCAGAT	840		
Qy	841	GCTCCGCTAGGCACACAGAGGATAAAAACAATCTTTATAGTATACCACTAAATTTTCGCTT	900		
Db	841	GCTCCGCTAGGCACACAGAGGATAAAAACAATCTTTATAGTATACCACTAAATTTTCGCTT	900		
Qy	901	AGTAACCTAGTGAATGTTTCAAGTCATGCTCCTGAGTCAAGAGTTGAGGAGACATTACATGT	960		

901	AGTAACCTAGTGAATGTTTCAAGTCATGCTCAGTCAAGAGTTGAGGAGACATTACAAATGT	960
961	GTAATGGAAACCAAGGAAAGTGAAACTTTTGATTAAGTGGGAGCTAGTGTATTTATATATT	1020
961	GTAATGGAAACCAAGGAAAGTGAAACTTTTGATTAAGTGGGAGCTAGTGTATTTATATATT	1020
1021	TAAATGATTTCTGACTCTATCATTTGGCCTCCAAACACACAGATTTGTTTCTTTCTGTTT	1080
1021	TAAATGATTTCTGACTCTATCATTTGGCCTCCAAACACACAGATTTGTTTCTTTCTGTTT	1080
1081	GTTTCTTTCACTATGGGATCTTCTGTGCCAGCAGTGCCTGCACACATAGAAAACAATC	1140
1081	GTTTCTTTCACTATGGGATCTTCTGTGCCAGCAGTGCCTGCACACATAGAAAACAATC	1140
1141	AATATTTCTGTAATTAATGATTAATAAATCAGAGAACCTTTCCCATCTCTGTTTGGATCTAT	1200
1141	AATATTTCTGTAATTAATGATTAATAAATCAGAGAACCTTTCCCATCTCTGTTTGGATCTAT	1200
1201	AGAACATCCAGAGTAAGTGATGAGGGCTCTGCATTTATATGCGCTTAAATTTAGATTAT	1260
1201	AGAACATCCAGAGTAAGTGATGAGGGCTCTGCATTTATATGCGCTTAAATTTAGATTAT	1260
1261	GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTGAAATATATAGTAAAAACA	1320
1261	GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTGAAATATATAGTAAAAACA	1320
1321	TGGTGGTCTTTAAAGAGATTAATAGATTAATGAAATCTCCATCTCAAAAATAATG	1380
1321	TGGTGGTCTTTAAAGAGATTAATAGATTAATGAAATCTCCATCTCAAAAATAATG	1380
1381	CATAAACTATTAAAGGAAATCAATCTCAGGGCTTTCAATGTTGTTCTACTTTT	1440
1381	CATAAACTATTAAAGGAAATCAATCTCAGGGCTTTCAATGTTGTTCTACTTTT	1440
1441	CATATATTTTACCATCTGCTGAAAGGAGTCAATCAAAAGGTTAAAGAAAGTGGAGGA	1500
1441	CATATATTTTACCATCTGCTGAAAGGAGTCAATCAAAAGGTTAAAGAAAGTGGAGGA	1500
1501	AAACCTCAGTAGAATTAATTAATGTTGTTGCAAGTAGAAAAGATTTCTCATCTCAA	1560
1501	AAACCTCAGTAGAATTAATTAATGTTGTTGCAAGTAGAAAAGATTTCTCATCTCAA	1560
1561	CCTTATGAGCAGGAGGAGGAGGCTGTTGAGAACCAATTTACTTTAGCAGAACACATAT	1620
1561	CCTTATGAGCAGGAGGAGGAGGCTGTTGAGAACCAATTTACTTTAGCAGAACACATAT	1620
1621	TTTAGACACTTTCCCTGCAATTAACCTGCAAAACAATATGTTTGCAAACTTTGTRGATCAAC	1680
1621	TTTAGACACTTTCCCTGCAATTAACCTGCAAAACAATATGTTTGCAAACTTTGTRGATCAAC	1680
1681	CTCCAAACACACACATTCAGGAGTTAAATATTTTTCATCAACATTTGGATTTTCTTTA	1740
1681	CTCCAAACACACACATTCAGGAGTTAAATATTTTTCATCAACATTTGGATTTTCTTTA	1740
1741	ACGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT	1800
1741	ACGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT	1800
1801	TTCTCCCTGTTTAAAGCAGAAACAAATTAAGCCATTTCACTCTCAAACTTCACTAATGA	1860
1801	TTCTCCCTGTTTAAAGCAGAAACAAATTAAGCCATTTCACTCTCAAACTTCACTAATGA	1860
1861	TCACATTTCTTCCAAAAGGAACTCTAGAAAGCCAAATGCCCCGAGTTAAGAACATCAAAA	1920
1861	TCACATTTCTTCCAAAAGGAACTCTAGAAAGCCAAATGCCCCGAGTTAAGAACATCAAAA	1920
1921	CTAACCTCTGAAGAACTTCCCAAGTGAAGACTCTGCTGCGACGACAAACATAAAAA	1980
1921	CTAACCTCTGAAGAACTTCCCAAGTGAAGACTCTGCTGCGACGACAAACATAAAAA	1980
1981	AAGAGAGAGGAAATCAATAGACACATTAATAAATGATAAAGGGATATCACCCAGCATCC	2040

Db 1981 AAGAGAGAAGATCAATATAGACACAAATATAAAGATATAAGGGGATATACACCGATCC 2040

Qy 2041 CACAGAAATACAAATACCAATCAGAGAAATACAAACACCTCTACGC 2088

Db 2041 CACAGAAATATAAAATACCAATCAGAGAAATACAAACACCTCTACGC 2088

RESULT 2

ARI146835

LOCUS

DEFINITION

Sequence 83 from patent US 6218529.

ACCESSION

ARI146835

VERSION

ARI146835.1 GI:15110024

KEYWORDS

Unknown.

SOURCE

Unknown.

ORGANISM

Unclassified.

REFERENCE

1 (bases 1 to 2087)

AUTHORS

An.G., O'Hara,S.Mark., Ralph,D. and Veltri,R.

TITLE

Biomarkers and targets for diagnosis, prognosis and management of prostate, breast and bladder cancer

JOURNAL

Patent: US 6218529-A 83 17-APR-2001;

FEATURES

Location/Qualifiers

1..2087

source

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Query Match 99.4%; Score 2076; DB 6; Length 2087;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2087; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 GACCTTAAATATATCGAGGTGGCTAAATTTGATGATATAATAATTTACAAAATTTCTTTCTA 60

Db 1 GACCTTAAATATATCGAGGTGGCTAAATTTGATGATATAATAATTTACAAAATTTCTTTCTA 60

Qy 61 TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCATGAGGCGCTTTAAAGAAC 120

Db 61 TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCATGAGGCGCTTTAAAGAAC 120

Qy 121 AGAAATATGAGGATATGACAAATATTATTTACAGATTCAGAAATTTGAGGCACA 180

Db 121 AGAAATATGAGGATATGACAAATATTATTTACAGATTCAGAAATTTGAGGCACA 180

Qy 181 GATTAAAGTAACTTCCCAAGCTACAGGCAATCTAGCTCCAGAAATCTGTCTTTACCAT 240

Db 181 GATTAAAGTAACTTCCCAAGCTACAGGCAATCTAGCTCCAGAAATCTGTCTTTACCAT 240

Qy 241 TCTGCTACAGGTATTTTCGAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCACAGAGT 300

Db 241 TCTGCTACAGGTATTTTCGAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCACAGAGT 300

Qy 301 TCATTGATTATTCATAGAACAGTACACCATGCAATTTCTCACCCCTTTGCAGACACCT 360

Db 301 TCATTGATTATTCATAGAACAGTACACCATGCAATTTCTCACCCCTTTGCAGACACCT 360

Qy 361 TGACCATGAAGGTTCCTCAATGAATGTTCCTCATTATCTTCAGAAAGCCATATTATCCA 420

Db 361 TGACCATGAAGGTTCCTCAATGAATGTTCCTCATTATCTTCAGAAAGCCATATTATCCA 420

Qy 421 CATTGACTTTGCAAGTAACTCAGACCCCTAGGTCTGGAATGCTGTCTCTCTACTTATCCA 480

Db 421 CATTGACTTTGCAAGTAACTCAGACCCCTAGGTCTGGAATGCTGTCTCTCTACTTATCCA 480

Qy 481 AAACATATACATCAGATATATAAATCTCAGCCCTCTCAGCCCTCTGCAAGCCCTTTCCAGAAA 540

Db 481 AAACATATACATCAGATATATAAATCTCAGCCCTCTCAGCCCTCTGCAAGCCCTTTCCAGAAA 540

Qy 541 ATAAAAATGGTTGAAAGGCAATTTCTGCTACCAATGACTGTTTAAAGCCAGCCCAAGTAAC 600

Db 541 ATAAAAATGGTTGAAAGGCAATTTCTGCTACCAATGACTGTTTAAAGCCAGCCCAAGTAAC 600

Qy 601 TGAACCAATTCCAACTTCAATTTACTTATGAAAGAAATTTGATGATGAGGTTATTTTC 660

Db 601 TGAACCAATTCCAACTTCAATTTACTTATGAAAGAAATTTGATGATGAGGTTATTTTC 660

Qy 661 AATTCTTAAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACCTCATAGATTATTTCT 720

Db 661 AATTCTTAAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACCTCATAGATTATTTCT 720

Qy 721 ATTATCTCAATTTAGTTTGTATTATCTCTAGTGGGCCATTTAAACATACCAATGTTGT 780

Db 721 ATTATCTCAATTTAGTTTGTATTATCTCTAGTGGGCCATTTAAACATACCAATGTTGT 780

Qy 781 TCTGCTCTCTCAATTTAGTCAATAAATCTTAACTACAGAGCAATTAGTAAGCCATGTCACAG 840

Db 781 TCTGCTCTCTCAATTTAGTCAATAAATCTTAACTACAGAGCAATTAGTAAGCCATGTCACAG 840

Qy 841 GCTCGCTAGGCACACAGAGGGATAAAAACAATCTATTATAGTATACCACTAAATTTTCGTT 900

Db 841 GCTCGCTAGGCACACAGAGGGATAAAAACAATCTATTATAGTATACCACTAAATTTTCGTT 900

Qy 901 AGTAACTAGTGAATTTGTTCAAGTCAAGCTGAGTCAAGAGTTGAGGAGACATTAACAATG 960

Db 901 AGTAACTAGTGAATTTGTTCAAGTCAAGCTGAGTCAAGAGTTGAGGAGACATTAACAATG 960

Qy 961 GTAAATGGAAACCAAGGAAGTGAAACTTTTGGATAAGTGGGACTAGTGTATTATATATT 1020

Db 961 GTAAATGGAAACCAAGGAAGTGAAACTTTTGGATAAGTGGGACTAGTGTATTATATATT 1020

Qy 1021 TAAATGATTTCTGACTCTATCATTTGGCCCTCAAAAACAGAGTTGTGTTTTCTTTGGTTTT 1080

Db 1021 TAAATGATTTCTGACTCTATCATTTGGCCCTCCAAAACAGAGTTGTGTTTTCTTTGGTTTT 1080

Qy 1081 GTTTCTTCACTATCGGATCTTCTGTGCCAGCACAGTGCCTGCACATAGAAAAACAATC 1140

Db 1081 GTTTCTTCACTATCGGATCTTCTGTGCCAGCACAGTGCCTGCACATAGAAAAACAATC 1140

Qy 1141 AATATTTCTGTAATTAATGATTTAAATAATCAGAGAACTTTCCCATCTGTGTTGGATCTAT 1200

Db 1141 AATATTTCTGTAATTAATGATTTAAATAATCAGAGAACTTTCCCATCTGTGTTGGATCTAT 1200

Qy 1201 AGAAATATCCAGAGTAAAGTGAAGGCGCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Db 1201 AGAAATATCCAGAGTAAAGTGAAGGCGCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260

Qy 1261 GTGAAAGGTTTAAAGACACCTTAGTAGAGTATTTTGAATATATAGTAAAACCTTTGAAA 1320

Db 1261 GTGAAAGGTTTAAAGACACCTTAGTAGAGTATTTTGAATATATAGTAAAACCTTTGAAA 1320

Qy 1321 TGGTGGTCTTTAAAGATATTAATAGATTAATGAAATCTCCATCTCAAAAATAATG 1380

Db 1321 TGGTGGTCTTTAAAGATATTAATAGATTAATGAAATCTCCATCTCAAAAATAATG 1380

Qy 1381 CATAAATCTATTATAAGGAAATACATCTCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Db 1381 CATAAATCTATTATAAGGAAATACATCTCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440

Qy 1441 CATATATTTTACCATCTGCTGAAGGAGTCAATCAAGGGTAAAGAAAGATGGAGGA 1500

Db 1441 CATATATTTTACCATCTGCTGAAGGAGTCAATCAAGGGTAAAGAAAGATGGAGGA 1500

Qy 1501 AAACCTCAGTAAAGATTTATTTAGTCTGTTTGCAGAGTAAAGAGATTTCTCATCTCAA 1560

Db 1501 AAACCTCAGTAAAGATTTATTTAGTCTGTTTGCAGAGTAAAGAGATTTCTCATCTCAA 1560

Qy 1561 CCTTATGAGCAGGAAGAGGAGGCTTTTGAAGAACCTTTACTTAGCAGAACCAATAT 1620

Db 1561 CCTTATGAGCAGGAAGAGGAGGCTTTTGAAGAACCTTTACTTAGCAGAACCAATAT 1620

Qy 1621 TTTAGACACTTCCCTGCAATTAATCTGCACAAACAATATGTTTGCAAACTGTTTGTATCAAC 1680

Db 1621 TTTAGACACTTCCCTGCAATTTAACTGCAACAACAATATGTTTGCAAACTGTTT-GATCAAC 1679

Qy 1681 CTCGAAACACGACATTCAGGAGTTAAATATTTTTTTCATCAACATTTGGATTTTTCCTTA 1740

```
Db 1680 CTCACAAACGACACATTCAGGAGTTAAATAATTTTTCATCAAAACATTTGGATTTTTCCTTA 1739
QY 1741 AGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTTAAGAGAGAT 1800
Db 1740 AGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTTAAGAGAGAT 1799
QY 1801 TTCTCCCTCTTATAAGCAGCAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAAATGA 1860
Db 1800 TTCTCCCTCTTATAAGCAGCAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAAATGA 1859
QY 1861 TCACATCTTTTCAAAAGAACTCTAGAGAGACAAATGGCCGAGTTAAGACATCAAAA 1920
Db 1860 TCACATCTTTTCAAAAGAACTCTAGAGAGACAAATGGCCGAGTTAAGACATCAAAA 1919
QY 1921 CTAACCATCTGAAGAACTTCCCAAGTGAAGACTCTGCTGCACGACACACATAAAA 1980
Db 1920 CTAACCATCTGAAGAACTTCCCAAGTGAAGACTCTGCTGCACGACACACATAAAA 1979
QY 1981 AAGAGAGAAGAAATCAATAGACACAAATAAAAAATGATAAGGGGATATCACCAACGATCC 2040
Db 1980 AAGAGAGAAGAAATCAATAGACACAAATAAAAAATGATAAGGGGATATCACCAACGATCC 2039
QY 2041 CACAGAATAACAACTACCATAGAGAACTACAAACCTCTAGC 2088
Db 2040 CACAGAATAACAACTACCATAGAGAACTACAAACCTCTAGC 2087

RESULT 3
AF189269
LOCUS Homo sapiens UC28 protein (UC28) mRNA linear PRI 22-MAR-2001
DEFINITION Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION AF189269
VERSION AF189269.1 GI:10441601
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2512)
AUTHORS An,G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
Wright,G.L. Jr. and Veltri,R.W.
Cloning and characterization of UROC28, a novel gene overexpressed
in prostate, breast, and bladder cancers
Cancer Res. 60 (24), 7014-7020 (2000)
MEDLINE 21028101
PUBMED 11156405
REFERENCE 2 (bases 1 to 2512)
AUTHORS An,G. and Veltri,R.W.
Direct Submission
TITLE Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
Parkway, Oklahoma City, OK 73104, USA
FEATURES
source
1..2512
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q23-q24"
/tissue_type="prostate cancer"
1..2512
/gene="UC28"
99..506
/gene="UC28"
/codon_start=1
/product="UC28 protein"
/protein_id="AAG17117.1"
/db_xref="GI:10441602"
/translation="MRAFLRNQKYMNMHNI IHIQIRKLRLHRLSPFRLPGILAPETV
LIPFCVKPRKKEKVKRSOKATEFIDYSIEQSHAILTPLQTHLTMKSGSMKCSLSLS
EAILFTLTQLNQTGLGECLLYLSKTHPQII"
polya_signal
2487..2492
/gene="UC28"

gene
CDS
901 AGTAACTAGTGAATTTGTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTAACAATGT 960
901 AGTAACTAGTGAATTTGTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTAACAATGT 960
961 GTAAATGGAAACCAAGGAAAGTGAAACCTTTGGATAGTGGGACTAGTGTATTTATATTT 1020
961 GTAAATGGAAACCAAGGAAAGTGAAACCTTTGGATAGTGGGACTAGTGTATTTATATTT 1020
```

```
Query Match
Best Local Similarity 93.9%; Score 1961.6; DB 9; Length 2512;
Matches 1968; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 GACCTTAAATATATCGAGGTGGCTAAATTTGATGATATAAATTTACAAAATTTATTTCTTCTA 60
Db 1 GACCTTAAATATATCGAGGTGGCTAAATTTGATGATATAAATTTACAAAATTTATTTCTTCTA 60
QY 61 TTGCTACAGAGCTACAATTCATTTTACAGTAGGCCACCATGAGGGCTTCTTTAAGGAACC 120
Db 61 TTGCTACAGAGCTACAATTTCAATTTTACAGTAGGCCACCATGAGGGCTTCTTTAAGGAACC 120
QY 121 AGAATATGAGGATATGACAAATATTTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
Db 121 AGAATATGAGGATATGACAAATATTTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
QY 181 GATTAAAGTAACTTCCCAAGGCTACCCAGGCACTTAGCTCCAGAAACTGTGCTTTTACCAT 240
Db 181 GATTAAAGTAACTTCCCAAGGCTACCCAGGCACTTAGCTCCAGAAACTGTGCTTTTACCAT 240
QY 241 TCTGCTACAAGTATTTTCGAAAAAAGAAAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 300
Db 241 TCTGCTACAAGTATTTTCGAAAAAAGAAAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 300
QY 301 TCATTGATTTTCCATAGAACAGTCACACCATGCAATTTCTCACACCTTTGAGGACACACT 360
Db 301 TCATTGATTTTCCATAGAACAGTCACACCATGCAATTTCTCACACCTTTGAGGACACACT 360
QY 361 TGACCATGAAAGGTTTCTCAATGAAATGTTCTCTCATTTTTCAGAAGCCATATTTATTC 420
Db 361 TGACCATGAAAGGTTTCTCAATGAAATGTTCTCTCATTTTTCAGAAGCCATATTTATTC 420
QY 421 CATTCAGCTTTGAGGTTAACTCAGACCCCTAGTCTGGAATGCTGCTTCTTCTACTTATCCA 480
Db 421 CATTCAGCTTTGAGGTTAACTCAGACCCCTAGTCTGGAATGCTGCTTCTTCTACTTATCCA 480
QY 481 AAATATATATCCACAGATCATATAAATCTTCAGCCCTGCTGCAAAAGCTTTTCAGAAAA 540
Db 481 AAATATATATCCACAGATCATATAAATCTTCAGCCCTGCTGCAAAAGCTTTTCAGAAAA 540
QY 541 ATAAAAATGGTTGAAAAGGCAATTCGTCTACCAATGCTTTTAAGCCACCAAGTAAC 600
Db 541 ATAAAAATGGTTGAAAAGGCAATTCGTCTACCAATGCTTTTAAGCCACCAAGTAAC 600
QY 601 TGAACCATTCCAATTCATTTTCTATGAAAAGAAATTTGATGATGATGATGATGATGATGAT 660
Db 601 TGAACCATTCCAATTCATTTTCTATGAAAAGAAATTTGATGATGATGATGATGATGATGAT 660
QY 661 AATTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTCAATCTCAATGATTTATTTCT 720
Db 661 AATTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTCAATCTCAATGATTTATTTCT 720
QY 721 ATTATCTCAATTTAGTTTGTATTATTCCTAGTGGCCATTTAAAAAATTTTTCGCTT 780
Db 721 ATTATCTCAATTTAGTTTGTATTATTCCTAGTGGCCATTTAAAAAATTTTTCGCTT 780
QY 781 TCTGCTCTCTCCATTTAGTCAATAAATACTTAACGAGCAATTTAGTAAAGCCATGTCAGAT 840
Db 781 TCTGCTCTCTCCATTTAGTCAATAAATACTTAACGAGCAATTTAGTAAAGCCATGTCAGAT 840
QY 841 GCTCGCTAGGCCACGAGGAGTAAAAACAATCTTATAGTATACCACTAAATTTTCGCTT 900
Db 841 GCTCGCTAGGCCACGAGGAGTAAAAACAATCTTATAGTATACCACTAAATTTTCGCTT 900
QY 901 AGTAACTAGTGAATTTGTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTAACAATGT 960
Db 901 AGTAACTAGTGAATTTGTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTAACAATGT 960
QY 961 GTAAATGGAAACCAAGGAAAGTGAAACCTTTGGATAGTGGGACTAGTGTATTTATATTT 1020
Db 961 GTAAATGGAAACCAAGGAAAGTGAAACCTTTGGATAGTGGGACTAGTGTATTTATATTT 1020
```

QY 1021 TAATTGATTTCTGACTATCATTTGGCTCCAAACACAGATTTGTTTCTTTGGTTT 1080
Db 1021 TAATTGATTTCTGACTATCATTTGGCTCCAAACACAGATTTGTTTCTTTGGTTT 1080
QY 1081 GTTTTCTTCACTATGGATCTTCTGTGCCAGACAGTGGCTGACACATAGAAAACAATC 1140
Db 1081 GTTTTCTTCACTATGGATCTTCTGTGCCAGACAGTGGCTGACACATAGAAAACAATC 1140
QY 1141 AATATTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCATCTGTTGGATCTAT 1200
Db 1141 AATATTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCATCTGTTGGATCTAT 1200
QY 1201 AGAACATCCAGAGTAAAGTATGAGGCTCTGCATTTATATGCTTAAATTAAGATTAT 1260
Db 1201 AGAACATCCAGAGTAAAGTATGAGGCTCTGCATTTATATGCTTAAATTAAGATTAT 1260
QY 1261 GTGAGAAAAGTTTAAAGACATTTAGTAGAGTGATTTTGAATAATAGTAAACACTTCGAAA 1320
Db 1261 GTGAGAAAAGTTTAAAGACATTTAGTAGAGTGATTTTGAATAATAGTAAACACTTCGAAA 1320
QY 1321 TGTGTGCTTTTAAAGATATTAATAGATAATATGAAATCTCCATCTCAAAAATAATG 1380
Db 1321 TGTGTGCTTTTAAAGATATTAATAGATAATATGAAATCTCCATCTCAAAAATAATG 1380
QY 1381 CATAACTATTTAAAGGAAATCACATCTCCAGGCTTTCATGCTTTGTTTCACTTTT 1440
Db 1381 CATAACTATTTAAAGGAAATCACATCTCCAGGCTTTCATGCTTTGTTTCACTTTT 1440
QY 1441 CATATATTTTACATCTGCTGAAGGAGTCAATCAAGGGTAAAGAAAGATGGAGGA 1500
Db 1441 CATATATTTTACATCTGCTGAAGGAGTCAATCAAGGGTAAAGAAAGATGGAGGA 1500
QY 1501 AAACCTCAGTAAAGTATATATTTAGTCTGTTTGCAGAGTAAAGAAAGATTTCTCATCTCAA 1560
Db 1501 AAACCTCAGTAAAGTATATATTTAGTCTGTTTGCAGAGTAAAGAAAGATTTCTCATCTCAA 1560
QY 1561 CCTTATGAGCAGAGGAGGAGGCTGTTTGAGAACATTTACTTAGCAGAACCATAT 1620
Db 1561 CCTTATGAGCAGAGGAGGAGGCTGTTTGAGAACATTTACTTAGCAGAACCATAT 1620
QY 1621 TTTAGACATTTCCCTGCTATTAACATGACAAACATATGTTTGCAAACTTTGTTGATCAAC 1680
Db 1621 TTTAGACATTTCCCTGCTATTAACATGACAAACATATGTTTGCAAACTTTGTTGATCAAC 1680
QY 1681 CTCCAAACAGCACATTCAGGAGTTAAATATTTTTCATCAAAATTTGGATTTTCTTTA 1740
Db 1681 CTCCAAACAGCACATTCAGGAGTTAAATATTTTTCATCAAAATTTGGATTTTCTTTA 1740
QY 1741 ACCTAGAGATTTGCTACAAATCTTCTGAAGGCTCAATGGCTTCAGGCTAAGAGAT 1800
Db 1741 ACCTAGAGATTTGCTACAAATCTTCTGAAGGCTCAATGGCTTCAGGCTAAGAGAT 1800
QY 1801 TTTCTCCTGTTTAAAGCAGACAAATTTAGGCAATTTCTACTCTCAAACTTCACTAATGA 1860
Db 1801 TTTCTCCTGTTTAAAGCAGACAAATTTAGGCAATTTCTACTCTCAAACTTCACTAATGA 1860
QY 1861 TCACATTTCTTCCAAAGGAACTCTAGAGACAAATGCCCGAGTTAAGAACATCAAAA 1920
Db 1861 TCACATTTCTTCCAAAGGAACTCTAGAGACAAATGCCCGAGTTAAGAACATCAAAA 1920
QY 1921 CTAACCATCTGAAGAACTTTCCCAAGTGAAGACTCTGCTGCACAGACACATAAA 1978
Db 1921 CTAACCATCTGAAGAACTTTCCCAAGTGAAGACTCTGCTGCACAGACACATAAA 1978

RESULT 4
LOCUS ARL146836 2505 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 85 from patent US 6218529.
ACCESSION ARL146836
VERSION ARL146836.1 GI:15110025
KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2505)
AUTHORS An, G., O'Hara, S., Mark, R., Ralph, D., and Veltri, R.
TITLE Biomarkers and targets for diagnosis, prognosis and management of prostate, breast and bladder cancer
JOURNAL Patent: US 6218529-A 85 17-APR-2001;
FEATURES Location/Qualifiers
 1..2505
 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN
Query Match 93.4%; Score 1950; DB 6; Length 2505;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 GACCTTAAATATATCGAGGTGCTAAATGATGTATATAATTTACAAAATTTCTTTCTTA 60
Db 1 GACCTTAAATATATCGAGGTGCTAAATGATGTATATAATTTACAAAATTTCTTTCTTA 60
QY 61 TTGCTACAGAGCTACAATTTCAATTTTACAGTAGGCCACCATGAGGCTTTCTTTAAGAAC 120
Db 61 TTGCTACAGAGCTACAATTTCAATTTTACAGTAGGCCACCATGAGGCTTTCTTTAAGAAC 120
QY 121 AGAATATGAGGATATGCAATATTTTCAATTTTACAGATCAGAAAATTTGAGGCACA 180
Db 121 AGAATATGAGGATATGCAATATTTTCAATTTTACAGATCAGAAAATTTGAGGCACA 180
QY 181 GATTAAAGTAACTTCCCAAGGCTACAGGCTTACAGGCTTCCAGGCTTCTAGCTTCTTACC 240
Db 181 GATTAAAGTAACTTCCCAAGGCTTACAGGCTTCCAGGCTTCTAGCTTCTTACC 240
QY 241 TCTGCTCAAGATTTTCGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 300
Db 241 TCTGCTCAAGATTTTCGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 300
QY 301 TCATTTGATTTTCCATAGAACAGTCAACCATGCAATTTCTCAACCTTTGACAGACACT 360
Db 301 TCATTTGATTTTCCATAGAACAGTCAACCATGCAATTTCTCAACCTTTGACAGACACT 360
QY 361 TGACCATGAAAGGTTCCCTCAATGAAATGTTCCCTCATTTCTCAGAGCCATATTATTTCA 420
Db 361 TGACCATGAAAGGTTCCCTCAATGAAATGTTCCCTCATTTCTCAGAGCCATATTATTTCA 420
QY 421 CATTTGACCTTTGAGTAACTCAGAGCCCTAGGCTGGAATGCTGTCTTCTCTACTTATCCA 480
Db 421 CATTTGACCTTTGAGTAACTCAGAGCCCTAGGCTGGAATGCTGTCTTCTCTACTTATCCA 480
QY 481 AAACATATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA 540
Db 481 AAACATATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA 540
QY 541 ATAAAAATGTTGAAAGGCAATTTCTGTTACCAATGCTGTTTAAAGCCAGGCAAGTAAAC 600
Db 541 ATAAAAATGTTGAAAGGCAATTTCTGTTACCAATGCTGTTTAAAGCCAGGCAAGTAAAC 600
QY 601 TGAACCATTTCCAACTTCAATTTTACATTTTGAAGAAATTTGATGATGTAGGAGTTATTTTC 660
Db 601 TGAACCATTTCCAACTTCAATTTTACATTTTGAAGAAATTTGATGATGTAGGAGTTATTTTC 660
QY 661 AATTCTAAATAACAAACCATTTGATTTCTTCAATCTTGAACCTCATAGATTTATTTCT 720
Db 661 AATTCTAAATAACAAACCATTTGATTTCTTCAATCTTGAACCTCATAGATTTATTTCT 720
QY 721 ATTATCTCAATTTAGTGTGTTTATTTTCTAGTGGGCAATTTAAACATCCACATGTGTT 780
Db 721 ATTATCTCAATTTAGTGTGTTTATTTTCTAGTGGGCAATTTAAACATCCACATGTGTT 780
QY 781 TCTGCTCTCCCAATTTAGTCAATACTAACTAACAGAGCAATTTAGTAAAGCCATGTGCCAGAT 840
Db 781 TCTGCTCTCCCAATTTAGTCAATACTAACTAACAGAGCAATTTAGTAAAGCCATGTGCCAGAT 840

JOURNAL

REFERENCE 2 (bases 1 to 186698)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (07-APR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Sep 1, 2000 this sequence version replaced gi:7523932.

ORIGIN

Query Match 92.0%; Score 1920.2; DB 2; Length 186698;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1961; Conservative 1; Mismatches 14; Indels 3; Gaps 3;
QY 1 GACCTTAATAATATATCGAGGTGGCTAAATTGATGTATATAAATTTACAAAATTTATTTCTTCTTA 60
Db 159156 GACCTTAATAATATATCGAGGTGGCTAAATTGATGTATATAAATTTACAAAATTTATTTCTTCTTA 159215
QY 61 TTGTACAGAGCTACAATTCAATTTTACAGTAGGCCACCATGAGGCGCTTCTTAAGGAACC 120
Db 159216 -TGCTACAGAGCTACAATTCAATTTTACAGTA-GTAAATATGAGGCGCTTCTTAAGGAACC 159273
QY 121 AGAAATATGAGGATATGCACAAATATTATTACATTTTACAGATCAGAAAATTTAGGCGACA 180
Db 159274 AGAAATATGAGGATATGCACAAATATTATTACATTTTACAGATCAGAAAATTTAGGCGACA 159333
QY 181 GATTAAAGTAACTTCCCAAGGCTACCGAGCATTTCTAGCTCCAGAAACTGTGCTCTTACCAT 240
Db 159334 GATTAAAGTAACTTCCCAAGGCTACCGAGCATTTCTAGCTCCAGAAACTGTGCTCTTACCAT 159393
QY 241 TCTGCTACAAGGTATTTTCGAAAAAGAAAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 300
Db 159394 TCTGCTACAAGGTATTTTCGAAAAAGAAAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 159453
QY 301 TCATTGATTATTTCCATAGACAGTCACACCATGCAATTTCTCACACCCCTTGAGACACACT 360
Db 159454 TCATTGATTATTTCCATAGACAGTCACACCATGCAATTTCTCACACCCCTTGAGACACACT 159513
QY 361 TGACCATGAAAGGTTTCTCAATGAAATGTTCTCTCATTTATCTTCAGAGCCATATTTATCA 420
Db 159514 TGACCATGAAAGGTTTCTCAATGAAATGTTCTCTCATTTATCTTCAGAGCCATATTTATCA 159573
QY 421 CATTGACCTTTGCGAGTTAACTCAGACCCCTAGTCTGGAATGCTGCTTCTCTACTTATCA 480
Db 159574 CATTGACCTTTGCGAGTTAACTCAGACCCCTAGTCTGGAATGCTGCTTCTCTACTTATCA 159633
QY 481 AAATATATATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGGCTTTTCAGAAAA 540
Db 159634 AAATATATATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGGCTTTTCAGAAAA 159693
QY 541 ATAAAAATGGTTGAAAAGGCAATTTCTGCTCAATGACTGTTTAAAGCCGCCCAAGTAAC 600
Db 159694 ATAAAAATGGTTGAAAAGGCAATTTCTGCTCAATGACTGTTTAAAGCCGCCCAAGTAAC 159753
QY 601 TGAACCATTCCAACTTCAATTTACTATGAAAAGAAATTTGATGATGTAGGAGGTTATTTTC 660
Db 159754 TGAACCATTCCAACTTCAATTTACTATGAAAAGAAATTTGATGATGTAGGAGGTTATTTTC 159813
QY 661 AATTCTAAAAATACAAACCCCATGTTGATCTTTCTCAATCTTCAACTCATAGATTATTTATCT 720
Db 159814 AATTCTAAAAATACAAACCCCATGTTGATCTTTCTCAATCTTCAACTCATAGATTATTTATCT 159873
QY 721 ATTATCTCAATTTAGTTTGTATTATCTTAGTGGCCATTAAAAAATACCACATGTGTT 780
Db 159874 ATTATCTCAATTTAGTTTGTATTATCTTAGTGGCCATTAAAAAATACCACATGTGTT 159933
QY 781 TCTGTCTCTCCATTAGTCAATAACTAAACGAGCAATTTAGTAAGCCATGTGCCAGAT 840
Db 159934 TCTGTCTCTCCATTAGTCAATAACTAAACGAGCAATTTAGTAAGCCATGTGCCAGAT 159993
QY 841 GCTCGCTAGCACCCAGAGGGATAAAAACAATCTTATAGTATACCACCTAAATTTTCGCTT 900
Db 159994 GCTCGCTAGCACCCAGAGGGATAAAAACAATCTTATAGTATACCACCTAAATTTTCGCTT 160053
QY 901 AGTAACTAGTCAAAATGTTCAAGTCATGCTCAGTCAAGAGTTGAGGAGACATTTACAATGT 960
Db 160054 AGTAACTAGTCAAAATGTTCAAGTCATGCTCAGTCAAGAGTTGAGGAGACATTTACAATGT 160113
QY 961 GTAATGGAAACCAAGGAAGTGAACTTTGGATAGTGGGAGCTAGTGTATTTATATTT 1020
Db 160114 GTAATGGAAACCAAGGAAGTGAACTTTGGATAGTGGGAGCTAGTGTATTTATATTT 160173

FEATURES

Source 1..186698
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="6"
/clone="RP11-133015"
1..4769
/note="assembly_name:Contig10"
4870..11657
/note="assembly_name:Contig11"
11758..28189
/note="assembly_name:Contig12"
28290..53341
/note="assembly_name:Contig13"
53442..102327
/note="assembly_name:Contig14"
102428..182781
/note="assembly_name:Contig15
clone_end:SP6
vector_side:right"
182882..186698
/note="assembly_name:Contig9"

QY	1021	TAATTGATTTCTGACTCTATCATTTGGCTCCAAACACAGATTTGTTTCTTTGTTT	1080
Db	160174	TAATTGATTTCTGACTCTATCATTTGGCTCCAAACACAGATTTGTTTCTTTGTTT	160233
QY	1081	GTTTTCTTCTACTATGGGATCTTCTGTGCCAGACAGTGGCTGACATAGAAAACAATC	1140
Db	160234	GTTTTCTTCTACTATGGGATCTTCTGTGCCAGACAGTGGCTGACATAGAAAACAATC	160293
QY	1141	-AATATTTGCTGAATTAATGATTAATAAATCAGAGAACTTTCCCATCTGTGTTGGATCTA	1199
Db	160294	AAATATTTGCTGAATTAATGATTAATAAATCAGAGAACTTTCCCATCTGTGTTGGATCTA	160353
QY	1200	TAGAACATCCAGATTAAGTATGAGGGCTCTGCATTTATATGCGCTTAAATTAAGATTA	1259
Db	160354	TAGAACATCCAGATTAAGTATGAGGGCTCTGCATTTATATGCGCTTAAATTAAGATTA	160413
QY	1260	TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGAATTTTGAATATAGTAAACACTTGGAA	1319
Db	160414	TGTGAGAAAAGTTTAAAGACACTTAGTAGAGTGAATTTTGAATATAGTAAACACTTGGAA	160473
QY	1320	ATGGTGGTCTTTAAAGAAATTAATATAGATAATATGAAAATCTCCATCTCAAAAATAAT	1379
Db	160474	ATGGTGGTCTTTAAAGAAATTAATATAGATAATATGAAAATCTCCATCTCAAAAATAAT	160533
QY	1380	GCATAACTATTTAAAGAAAATCAGATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT	1439
Db	160534	GCATAACTATTTAAAGAAAATCAGATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT	160593
QY	1440	TCATATATTTTACCATCTCTGAAGCAGTCAATCAAGGGTAAAGAAAGATGGAGG	1499
Db	160594	TCATATATTTTACCATCTCTGAAGCAGTCAATCAAGGGTAAAGAAAGATGGAGG	160653
QY	1500	AAAACCTCAGTAAGATTAATATAGTCTGTGTTGCAAAAGTAGAAAAGATTTCTCATCTCA	1559
Db	160654	AAAACCTCAGTAAGATTAATATAGTCTGTGTTGCAAAAGTAGAAAAGATTTCTCATCTCA	160713
QY	1560	ACCTTATGAGCAGGAAGAGGAGGCTGTTGAGAACCATTTACTTAGAGAACACATATA	1619
Db	160714	ACCTTATGAGCAGGAAGAGGAGGCTGTTGAGAACCATTTACTTAGAGAACACATATA	160773
QY	1620	TTTTAGACACTTCCCTGCTTAATCTCAGAAAATATGTTTGCACAACTTTGTTGATCAA	1679
Db	160774	TTTTAGACACTTCCCTGCTTAATCTCAGAAAATATGTTTGCACAACTTTGTTGATCAA	160833
QY	1680	CTTCCAAACACGACACATTCAGGAGTTAAATATTTTTCATCAACATTTGATTTTCCCTT	1739
Db	160834	CTTCCAAACACGACACATTCAGGAGTTAAATATTTTTCATCAACATTTGATTTTCCCTT	160893
QY	1740	AACGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAAGAG	1799
Db	160894	AACGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAAGAG	160953
QY	1800	TTTCTCCCTGTTATAGCAGCAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATG	1859
Db	160954	TTTCTCCCTGTTATAGCAGCAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATG	161013
QY	1860	ATCACATTTCTTTCCAAAAGGAACTCTAGAGAGCAAAATGCCCGAGTTAAGAACATCAAA	1919
Db	161014	ATCACATTTCTTTCCAAAAGGAACTCTAGAGAGCAAAATGCCCGAGTTAAGAACATCAAA	161073
QY	1920	ACTAACATCTGAAGAACTTCCCAAGTGTAGACTCTGCTGCGACAGACACATATAA	1978
Db	161074	ACTAACATCTGAAGAACTTCCCAAGTGTAGACTCTGCTGCGACATTAACCAATACCAGAG	161132

RESULT 7
BC069109
LOCUS BC069109 2496 bp mRNA linear PRI 23-JUN-2004
DEFINITION Homo sapiens prostate and breast cancer overexpressed 1, mRNA (cDNA
clone IMAGE:7216926).
ACCESSION BC069109
VERSION BC069109.1 GI:46575743

KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL
PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK
COMMENT

FEATURES
source

ORIGIN

Query Match 91.9%; Score 1919.2; DB 9; Length 2496;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1960; Conservative 1; Mismatches 14; Indels 3; Gaps 3;

QY 2 ACCTTAAATATATCGAGTGGCTTAATTCATGTATATTAATTTACAAAATTTCTTCTAT 61
Db 1 ACCTTAAATATATCGAGTGGCTTAATTCATGTATATTAATTTACAAAATTTCTTCTA- 59
QY 62 TGCTACAGAGCTACAAATTCATTTACAGTAGCCACCATGAGGCGCTTCTTAAGGAACCA 121
Db 60 TGCTACAGAGCTACAAATTCATTTACAGTA-CTAAATATGAGGCGCTTCTTAAGGAACCA 118

Homo sapiens (human)
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2496)
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L.,
Scheetz T.E., Brownstein M.J., Udin T.B., Toshiyuki S.,
Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J.,
Armsom R.D., Muliyil S.J., Bosak S.A., McEwan P.J.,
McKernan K.J., Malek J.A., Gunaratne P.H., Richards S.,
Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettner M., Madan A., Young A.C., Shevchenko Y.,
Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D.,
Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S., Krzywinski M.I., Skalska U., Small D.E.,
Schnerch A., Schein J.E., Jones S.J. and Marra M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2496)
Strausberg R.
Direct Submission
Submitted (16-APR-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-roman@nih.gov
Tissue Procurement: Anup Madan, University of Iowa
cDNA Library Preparation: Anup Madan, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Neurogenomics Research Lab,
200 B EMBR, University of Iowa, Iowa City, IA-52242
anup-madan@uiowa.edu
Jessica Fahey, Tim Nelson, Jae Goon Yoon and Anup Madan
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: Plate: Row: Column: 0.
Location/Qualifiers
1..2496
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7216926"
/tissue_type="Synthetic constructs"
/clone_lib="NIH MGC_242"
/lab_host="XL10"
/note="Vector: pUC19"


```
Query Match      13.2%; Score 275.4; DB 6; Length 757;
Best Local Similarity 96.2%; Pred. No. 3e-42;
Matches: 282; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1686 ACAACGACATTCAGGAGTTAAATATTTTCATCAAAACATGGATTTTTCCTTAACGCT 1745
      |||
      1 ACACGACACATTCAGGAGTTAAATATTTTCATCAAAACATGGATTTTTCCTTAACGCT 60

QY 1746 AGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGCTTCAGGCTAAGAGAGATTTCCTC 1805
      |||
      61 AGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGCTTCAGGCTAAGAGAGATTTCCTC 120

QY 1806 CCTGTTATAGCAGCAGCAAGCAAAATTAGCCATTTCATCTCCTCAAACTTCATATGATCACA 1865
      |||
      121 CCTGTTATAGCAGCAGCAGCAAAATTAGCCATTTCATCTCCTCAAACTTCATATGATCACA 180

QY 1866 TTTCTTTCCAAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGAACATCAAAACCTAAC 1925
      |||
      181 TTTCTTTCCAAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGAACATCAAAACCTAAC 240

QY 1926 CATCTGAAGAAACTTCCCAAGTGTAAAGACTCTGCTCGCAGCAACACATATAA 1978
      |||
      241 CATCTGAAGAAACTTCCCAAGTGTAAAGACTCTGCTCGCAGCAACATTAACCGAGA 293

RESULT 9
AC131318      230685 bp      DNA      linear      HTG 20-FEB-2004
LOCUS      Mus musculus chromosome 10 clone RP23-367D2 map 10, *** SEQUENCING
DEFINITION      IN PROGRESS ***, 8 unordered pieces.
ACCESSION      AC131318
VERSION      AC131318.4 GI:42716133
KEYWORDS      HTG; HTGS PHASE1; HTGS FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 230685)
Birken,B., Nuebaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Navlor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 230685)
Birken,B., Nuebaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cooke,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,

```

```

Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Navlor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (20-FEB-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 20, 2004 this sequence version replaced gi:29150479.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26968
Center clone name: 367_D_2
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 47907: contig of 47907 bp in length
* 47908: gap of 100 bp
* 48008 74241: contig of 26234 bp in length
* 74242 74341: gap of 100 bp
* 74342 89037: contig of 14696 bp in length
* 89038 89137: gap of 100 bp
* 89138 118227: contig of 29090 bp in length
* 118228 118327: gap of 100 bp
* 118328 158787: contig of 40460 bp in length
* 158788 158887: gap of 100 bp
* 158888 171882: contig of 12995 bp in length
* 171883 171982: gap of 100 bp
* 171983 196741: contig of 24759 bp in length
* 196742 196841: gap of 100 bp
* 196842 230685: contig of 33844 bp in length.

Location/Qualifiers
1. 230685
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10090"
   /chromosome="10"
   /map="10"
   /clone="RP23-367D2"
   /clone_lib="RPCI-23 Female Mouse BAC"

FEATURES
source

ORIGIN

Query Match      6.2%; Score 129.2; DB 2; Length 230685;
Best Local Similarity 59.3%; Pred. No. 2.8e-15;
Matches 416; Conservative 1; Mismatches 244; Indels 41; Gaps 10;

QY 1194 GATCTATAGAACATCCAGAGTAAGTATGATGAGGGCTCTGCTTATATATCGCTTAATTA 1253
      |||
      125132 GAATTTCTAGAACATTCACAGTAATGCTGAG-----AGCTTACATCTGTGCTTAACCTTC 125185

QY 1254 AGATTATGTGAGAAAAGTTTAAAG---ACACTTGTAGTAGTGTATTTGAAATATAGTAAA 1310
      |||
      125186 TGATCATAGTGGAGAGATTATTAGGTAGTAGTGTCTCAAGTATGTCCAAATGGTACAGT 125245
```

```
QY 1311 CACTTGGAAATGGTGGTCTTTAAAGGATATTAATAGATATAATGAAATCTCCATCTC 1370
Db 125246 ACTTAGACATGGTGATACCTTATGAAAGAGAGTTAATAGATATAATGAAACCT-CATCTC 125304
QY 1371 AAAAATAATGCATAAATACTATTAAAGGAAATACATCTCCAGGCTTTCAATGTTTGTTC 1430
Db 125305 ACANAAGAGTGAATAACGAAAGAAATAATTTACAGTTATGAGACTTCTGCAATGTTGCTTC 125364
QY 1431 ATTACTTTTTTCATATATTTTACCATCTGCTGAAGGCAGTCATATCAAAAGGGTAAAGAAA 1490
Db 125365 TTCT-----TTAATAATCATCTGCCTAAGGTATGTTATTTCAAAGGGTAAA-AGA 125411
QY 1491 GATGGAGGAAACCTCAGTAAGAATTATATTAGTCTGTGTTGCAAGTAGAAAAAGATTCT 1550
Db 125412 GAAAGAAGGAAATAATAATTAGTTATGTTATAAATGTGTGCCAGGTAGAAAATCTTAT 125471
QY 1551 CATCACTCAACCTTATGACAGAGGAGGAGGCTGTTTGAGAACCACTTACTTAGCAG 1610
Db 125472 C-TCAGTCAACATTATGACAGCAGGAA---GGAAGAGCTGCCCTAAAAACAATAATTACTTAGGAG 125527
QY 1611 AACCACA-----TATTTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTT 1661
Db 125528 AACCAACACAGGTTGGTAGATATGCTGTAAGTGTGTTAGACTCAACCCCAACACAGCA 125587
QY 1662 GCAAACTTGTTRGAT-CAACCTCCAAACAGCACATTCAGGAGTTAAATATTTTTTCATC 1720
Db 125588 GTTAGGAGTTATATACAATTTCCAAACAGCAGTAGCTGTAGACTCATACATTTTTTCATC 125647
QY 1721 AAACATTGATTTTCTTAAAGCTAGAGATTGCTACAATCTTCTGAAGGCTCTCAATG 1780
Db 125648 AAGCATTTGATTTCTCTTAGCGCCAAAGACTATTATAAATCTTCCAAAGAATCGCTATG 125707
QY 1781 GCTTCAGGCTAAGAAG---AGATTCTCCCTGTTATAGCAGCAGACAAATTAAGCCAT 1836
Db 125708 GUTTAGCGTAAGAAGGGCTTGCTTGCCCTTGATATAGGAGGTAAACAATCAGTCAT 125767
QY 1837 TTCACCTCTCAAACTTCACATAATGATCACATCTCTTCCAAAAG 1878
Db 125768 TTTAGCCTCAAGCTTGTCTCATGATCATCTTTTCCATAGAAG 125809

RESULT 10
AC026382
LOCUS
DEFINITION Mus musculus clone RP23-246G10, WORKING DRAFT SEQUENCE, 19
unordered pieces.
ACCESSION AC026382
VERSION AC026382.8 GI:16118081
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 167998)
Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuon,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Louleaged,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
Peretz,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
REFERENCE
2 (bases 1 to 167998)
Worley,K.C.
```

TITLE
JOURNAL

COMMENT

```
Direct Submission
Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 14, 2001 this sequence version replaced gi:11094624.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MADA
Center clone name: RP23-246G10
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-Primer Bodipy: 96% of reads
Chemistry: Dye-terminator Big Dye: 4% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 159501 bases at least Q40
Consensus quality: 166426 bases at least Q30
Consensus quality: 169264 bases at least Q20
Estimated insert size: 169620; sum-of-contigs estimation
Estimated insert size: 152000; agarose-fp estimation
Quality coverage: 4.5x in Q20 bases; agarose-fp estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 19 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
be preserved.
* 1 29927: contig of 29927 bp in length
* 29928 30027: gap of unknown length
* 30028 56267: contig of 26240 bp in length
* 56267 56367: gap of unknown length
* 56367 69896: contig of 13529 bp in length
* 69896 69996: gap of unknown length
* 69996 83796: contig of 13800 bp in length
* 83796 83896: gap of unknown length
* 83896 93698: contig of 9802 bp in length
* 93698 93798: gap of unknown length
* 93798 102657: contig of 8859 bp in length
* 102657 102757: gap of unknown length
* 102758 113241: contig of 10484 bp in length
* 113241 113341: gap of unknown length
* 113341 120111: contig of 6770 bp in length
* 120111 120211: gap of unknown length
* 120211 128008: contig of 8597 bp in length
* 128008 128908: gap of unknown length
* 128908 134501: contig of 5593 bp in length
* 134501 134601: gap of unknown length
* 134601 140411: contig of 5810 bp in length
* 140411 140511: gap of unknown length
* 140511 144731: contig of 4220 bp in length
* 144731 144831: gap of unknown length
* 144831 149476: contig of 4645 bp in length
* 149476 149576: gap of unknown length
* 149576 153295: contig of 3719 bp in length
* 153295 153395: gap of unknown length
* 153395 156120: contig of 2725 bp in length
* 156120 156220: gap of unknown length
* 156220 159545: contig of 3325 bp in length
* 159545 159645: gap of unknown length
* 159645 162911: contig of 3266 bp in length
* 162911 163011: gap of unknown length
* 163011 165637: contig of 2626 bp in length
* 165637 165737: gap of unknown length
* 165737 167998: contig of 2261 bp in length.
* 167998
```

TITLE
JOURNAL
REFERENCE
AUTHORS

FEATURES
source Location/Qualifiers
1.167998 /organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-246G10"

ORIGIN

Query Match 5.9%; Score 123.2; DB 2; Length 167998;
Best Local Similarity 58.7%; Pred. No. 4.3e-14;
Matches 412; Conservative 1; Mismatches 249; Indels 40; Gaps 10;

QY 1194 GATCTATAGACATCCAGAGTATGATGAGGCGCTCTGCTTATATATGCGCTTAATTA 1253
Db 51613 GAATCTCAGAACATTCACAGTAATGCTGAG-----AGCTTACATCTGTGCTTAACTTC 51666

QY 1254 AGATTATGTGAGAAAGTTTAAAG---ACACTTAGTAGAGTGATTTTGAATATAGTAAA 1310
Db 51667 TGATCATAGTGGGAGAGTTATTAGGTAGTAGTCTCAAGTATGTCCNATGGTACAGT 51726

QY 1311 CACTTGAATAGTGGTCTTTTAAAGATATTAATAGATAATGAAATCTCCATCTC 1370
Db 51727 ACCTAGACATGGTGATCTTATGAAAGAGGTTAATAGATAATATGAAACCT-CATCTC 51785

QY 1371 ARAAATATGCATAAATCTTTTAAAGGAAATCACATCTCCAGGCTTCAATGTTTGTTC 1430
Db 51786 ACAAGAGTGAATAACGAAAGAAATAATTCAGTATGAGCTTCTGCAATGTTGCTTC 51845

QY 1431 ATTACTTTTTTCATATATTTTTTACCATCTGCTGAAGCAGTCATCAAAAGGGTAAAGAAA 1490
Db 51846 TTCT-----TTATAATCATCTGCTAAGGTAGTTATTTCAAGGGTAAATGA 51893

QY 1491 GATGGG-AGAAACTCAGTAAGAAATATATATAGTCTGTTTGCAGAGTAAAGAAAGATTC 1549
Db 51894 GAAAGTAAGGAAATATAATTAGTATGTATAAATGTGTGCGCAGGTAGAAATACTTTA 51953

QY 1550 TCATCACTCAACTTATGACGAGAGGAGGAGGCTGTTTGAACCACTTACTTAGCA 1609
Db 51954 TC-TCAGTCAACATTAGCAGCAGGAA---GGAAAGCTGCCCTAAAGAAACAAATTA 52009

QY 1610 GAAC-CACATATTTTAGACATCTTCCCTGCAATTAACCTGCACAAACAATATGTTGCAAACT 1668
Db 52010 GAACACACACGGTTGGTAGATGTCTGTAGGTTGTAGACTCAACCCCAACACAGCA 52069

QY 1669 TGTRG-----ATCAACTCCAAACGACACATTCAGGAGTTAAATATTTTTCATC 1720
Db 52070 GTTAAGGAGTATATACAATTTCCACAGCAGTACGCTGTAGACTCATACATTTTTCATC 52129

QY 1721 AATCATGATTTTCTTCAACGCTAGAGATTGCTACAAATCTTCTGAAGGGTCTCAATG 1780
Db 52130 AAGCATTTGATTTCTCTTAGCGCCAAAGACTATTATAAATCTTCCAAAGAAATCGCTATG 52189

QY 1781 GCTTCAGGCTAAGAAGAGA---TTTCTCCCTGTATATAGCAGCAAGACAAATAGCCAT 1836
Db 52190 GCTTAGGTAAGAGGGCTTGCTTGGCCCTTGATATATAGGAGGTAACCACTCACTCAT 52249

QY 1837 TTCACCTCAAACTTCACATATGATCACAATCTTTTCCAAAAG 1878
Db 52250 TTTAGCGCTCAAGCTTTGCTCATGATCATCTTCTTCCATAGAAG 52291

RESULT 11
AC114782_1
WPCOMMENT
Sequence split into 5 fragments LOCUS AC114782 Accession AC114782
Fragment Name Begin End
AC114782_0 1 110000
AC114782_1 100001 210000
AC114782_2 200001 310000
AC114782_3 300001 410000
AC114782_4 400001 474678
Continuation (2 of 5) of AC114782 from base 100001 (AC114782 Homo sapiens chromosome UNK

Query Match 5.8%; Score 122; DB 2; Length 110000;
Best Local Similarity 89.7%; Pred. No. 8.1e-14;
Matches 131; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1943 CAAGTGTAAAGACTCTGCTGCACGACACACACATAAAGAGAGAGAGAAATCAAAATAGAC 2002
Db 54002 CAAATTTGATAGACCGCTAGCAGATGACAAGAAAAAGAGAGAGAAATCAATAGAC 54061

QY 2003 ACAATAAAAAATGATAAAGGGGATATCACACCGATCCACAGAAATACAAACTACCATC 2062
Db 54062 ACAATAAAAAATGATAAAGGGGATATCACACCGATCCACAGAAATACAAACTACCATC 54121

QY 2063 AGAGAAATACATAAAGACCTCTACGC 2088
Db 54122 AGAGAAATACATAAAGACCTCTACGC 54147

RESULT 12
AC093897
LOCUS
DEFINITION Homo sapiens BAC clone RP11-721G13 from 4, complete sequence.
ACCESSION AC093897 AC067901
VERSION AC093897.3 GI:16304304
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163271)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 163271)
AUTHORS Van Brunt, A. and Kozlowski, A.
TITLE The sequence of Homo sapiens BAC clone RP11-721G13
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 163271)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 163271)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 163271)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 21, 2001 this sequence version replaced gi:15825639.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0721G13
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate


```
Query Match          5.8%; Score 122; DB 9; Length 163271;
Best Local Similarity 89.7%; Pred. No. 7.3e-14;
Matches 131; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1943 CAAAGTGTAAAGACTCTGCTGCAGCACACACATATAAAAAAGAGAGAGAGAAATCAAAATAGAC 2002
    |||
DB 40437 CAAAATGTATAGACCCGCTAGCAGATGACANAGAAAAAAGAGAGAAATCAAAATAGAC 40496

QY 2003 ACAATAAAAAATGATAAAGGGGATATCACACCGATCCACACAGAAATACAAATCAATC 2062
    |||
DB 40497 ACAATAAAAAATGATAAAGGGGATATCACACCGATCCACACAGAAATACAAATCAATC 40556

QY 2063 AGAGATACTACAAACACCTCTACGC 2088
    |||
DB 40557 AGAGATACTACAAACACCTCTACGC 40582

RESULT 13
AC117430/c
LOCUS
DEFINITION
Homo sapiens 3 BAC RP11-119D18 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
ACCESSION AC117430
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 154941)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anarutunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbiera,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroil,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,P., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louleghed,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nickerson,E., Nwoketwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,D., Peters,L.,
Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E.,
Sonaika,T., Sparks,A., Stanley,H., Stone,H., Tansey,J., Taylor,C.,
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
Warren,R., Washington,C., Watlington,S., Williams,G.,
Williamson,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,Y.,
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
Gibbs,R.
Direct Submission
```

```
JOURNAL REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

```
Unpublished
2 (bases 1 to 154941)
Worley,K.C.
Direct Submission
Submitted (10-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 154941)
Worley,K.C.
Direct Submission
Submitted (29-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 154941)
Worley,K.C.
Direct Submission
Submitted (30-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 154941)
Worley,K.C.
Direct Submission
Submitted (27-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 30, 2002 this sequence version replaced gi:20279244.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.
Location/Qualifiers
1..154941
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-119D18"
1..2000
/note="overlaps bases 113280..115279 of clone AC112506"
```

```
/function="clone overlap"
repeat_region 1. .538
/rpt_family="L1M3b"
repeat_region 910. .1310
/rpt_family="L2"
repeat_region 2848. .2899
/rpt_family="TA)n"
repeat_region 3958. .3979
/rpt_family="AT_rich"
repeat_region complement(5020. .5227)
/rpt_family="MIR"
repeat_region 6721. .6843
/rpt_family="LTR48"
repeat_region complement(12729. .13025)
/rpt_family="AluY"
repeat_region 13781. .13843
/rpt_family="TA)n"
repeat_region complement(14404. .14713)
/rpt_family="AluSg"
repeat_region 14716. .14744
/rpt_family="AT_rich"
repeat_region 14807. .14861
/rpt_family="MIR"
repeat_region 14862. .15148
/rpt_family="AluSc"
repeat_region 15149. .15255
/rpt_family="MIR"
repeat_region 15680. .16066
/rpt_family="L1MA2"
repeat_region 16915. .17029
/rpt_family="FLAM_C"
repeat_region 17052. .17097
/rpt_family="FLAM_A"
repeat_region 17098. .17452
/rpt_family="THE1B"
repeat_region 17453. .17538
/rpt_family="FLAM_A"
repeat_region 17760. .17790
/rpt_family="AT_rich"
repeat_region 17943. .17975
/rpt_family="T-rich"
repeat_region complement(17977. .18298)
/rpt_family="AluJo"
repeat_region 18947. .18967
/rpt_family="(TTTA)n"
repeat_region complement(18970. .24273)
/rpt_family="L1PB1"
repeat_region complement(24298. .25342)
/rpt_family="L1PB1"
repeat_region 25434. .25475
/rpt_family="T-rich"
repeat_region 26099. .26122
/rpt_family="AT_rich"
repeat_region complement(26363. .26501)
/rpt_family="L1MA2"
repeat_region 26674. .26985
/rpt_family="AluSg"
repeat_region 27131. .27200
/rpt_family="AT_rich"
repeat_region complement(27385. .28146)
/rpt_family="L1MA4"
repeat_region 28219. .28301
/rpt_family="A-rich"
repeat_region complement(28751. .29047)
/rpt_family="AluJB"
repeat_region complement(29048. .29125)
/rpt_family="MIR"
repeat_region 29505. .29812
/rpt_family="LTR16A"
repeat_region complement(30217. .30329)
/rpt_family="MER8"
repeat_region complement(30330. .30426)
/rpt_family="HERV16"

repeat_region 30572. .30742
Query Match 5.8%; Score 121; DB 9; Length 154941;
Best Local Similarity 78.4%; Fred. No. 1.1e-13;
Matches 145; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 1904 AGTTAAGAACATCAAACTTAACCATCTCGAAGAACTTCCCAAGTGTGAAGACTCTGCCTGC 1963
Db 147147 AATTATGATCCAGGAGCTAGTTTTTTGAAGGATCAACAAATTTGATAAATGCTAGC 147088
QY 1964 AGCAACACATATAAAAAAGAGAGAGAAATCAATAGACACAATAAAAAATGATAAAGG 2023
Db 147087 AAGACTAATAAGAAAAAAGAGAGAGAAATCAATAGACACAATAAAAAATGATAAAGG 147028
QY 2024 GATATCACCGGATCCACAGAAATACAACTACCATCAGAGATACATAAACACCTC 2083
Db 147027 GATATCACCGGATCCACAGAAATACAACTACCATCAGAGATACATAAACACCTC 146968
QY 2084 TACGC 2088
Db 146967 TACGC 146963

RESULT 14
AC053487
LOCUS Homo sapiens chromosome 3 clone RP11-119D18, WORKING DRAFT
DEFINITION AC053487
ACCESSION AC053487.3 GI:9958145
VERSION AC053487.3
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155722)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155722)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 1, 2000 this sequence version replaced gi:9795672.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0119D18
----- Summary Statistics -----
Sequencing vector: M13; 76%
Sequencing vector: plasmid; 24%
Chemistry: Dye-terminator Big Dye; 26% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 154612 bases at least Q40
Consensus quality: 154798 bases at least Q30
Consensus quality: 154903 bases at least Q20
Insert size: 132000; agarose-fp
Insert size: 154934; sum-of-contigs
Quality coverage: 8.60 in Q20 bases; agarose-fp
Quality coverage: 7.32 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```

* as soon as it is available and the accession number will
* be preserved.

* 1 154934: contig of 154934 bp in length
* 154935 155034: gap of unknown length
* 155035 155722: contig of 688 bp in length.

FEATURES

source
1..155722
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-119D18"
misc_feature 1..154934
/note="assembly_name:Contig43
vector_end:SP6
vector_side:right"
misc_feature 155035..155722
/note="assembly_name:Contig15"

ORIGIN

Query Match 5.8%; Score 121; DB 2; Length 155722;
Best Local Similarity 78.4%; Pred. No. 1.1e-13;
Matches 145; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 1904 AGTTAGACATCAAACTAACCATCTGAAGAACTTCCCAAGTGTAACTCTGCCTGC 1963
DB 7795 AATTAATGAATCAGGAGCTAGTTTTTGAAGAGATCAACAAATTTGATAAACTGTAGC 7854
QY 1964 ACAGCAACACATAAAAAAGAGAGAGAGATCAATAGACACATAAAAAATGATAAGGG 2023
DB 7855 AGACTTAATAAGAAAAAGAGAGAGAGATCAATAGACACATAAAAAATGATAAGGG 7914
QY 2024 GATATCACCACCGATCCACAGAAATACAAATACCATCAGAGAATPACTACAAACCTTC 2083
DB 7915 GATATCACCACCGATCCACAGAAATACAAATACCATCAGAGAATPACTACAAACCTTC 7974
QY 2084 TAGCG 2088
DB 7975 TAGCG 7979

RESULT 15

HUMRETELAS
LOCUS 180388 bp DNA linear PRI 23-NOV-1994
DEFINITION Human retinoblastoma susceptibility gene exons 1-27, complete cds.
ACCESSION L11910
VERSION L11910.1 GI:292420
KEYWORDS nuclear protein; recessive oncogene; retinoblastoma gene;
retinoblastoma protein; retinoblastoma susceptibility; tumor
suppressor gene.

SOURCE

ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Friend,S.H., Bernards,R., Rogelj,S., Weinberg,R.A., Rapaport,J.M.,
Albert,D.M. and Dryja,T.P.

TITLE A human DNA segment with properties of the gene that predisposes to
retinoblastoma and osteosarcoma

JOURNAL

MEDLINE Nature 323 (6089), 643-646 (1986)
PUBMED 87039336
REFERENCE 2877398

AUTHORS

Friend,S.H., Horowitz,J.M., Gerber,M.R., Wang,X.F., Bogenmann,E.,
Li,F.P. and Weinberg,R.A.

TITLE Deletions of a DNA sequence in retinoblastomas and mesenchymal
tumors: organization of the sequence and its encoded protein

JOURNAL

MEDLINE Proc. Natl. Acad. Sci. U.S.A. 84 (24), 9059-9063 (1987)
PUBMED 88097427
REFERENCE 3480530

AUTHORS

Lee,W.H., Bookstein,R., Hong,F., Young,L.J., Shew,J.Y. and Lee,E.Y.

TITLE Human retinoblastoma susceptibility gene: cloning, identification,

JOURNAL

MEDLINE Science 235 (4794), 1394-1399 (1987)
PUBMED 87149066

REFERENCE

AUTHORS 4 (sites)

TITLE

McGee,T.L., Yandell,D.W. and Dryja,T.P.
Structure and partial genomic sequence of the human retinoblastoma
susceptibility gene
Gene 80 (1), 119-128 (1989)

JOURNAL

MEDLINE Nature 353 (6339), 83-86 (1991)
PUBMED 91351319

REFERENCE

AUTHORS 5 (sites)

TITLE

Sakai,T., Ohtani,N., McGee,T.L., Robbins,P.D. and Dryja,T.P.
Oncogenic germ-line mutations in Sp1 and ATF sites in the human
retinoblastoma gene

JOURNAL

MEDLINE Nature 353 (6339), 83-86 (1991)
PUBMED 91351319

REFERENCE

AUTHORS 6 (bases 1 to 180388)

TITLE

Toguchida,J., McGee,T.L., Paterson,J.C., Eagle,J.R., Tucker,S.,
Yandell,D.W. and Dryja,T.P.
Complete genomic sequence of the human retinoblastoma
susceptibility gene

JOURNAL

MEDLINE Genomics 17 (3), 535-543 (1993)
PUBMED 94063891

COMMENT

Original source text: Homo sapiens DNA.

FEATURES

source
1..180388
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1858..1863
/bound_moiety="Sp1"
/evidence="experimental"
1866..1872
/bound_moiety="ATF"
/evidence="experimental"
Join(2060..2196,5424..5550,39446..39561,41926..42045,
44668..44706,45799..45866,56853..56963,59651..59793,
61730..61807,64330..64439,65364..65441,70242..70329,
73753..73869,76430..76486,76889..76920,77001..77077,
78083..78279,149998..150116,153208..153353,156693..156838,
160730..160834,161997..162110,162204..162367,
170372..170402,173707..173849,174361..174410,
177005..177078)
/codon_start=1
/product="retinoblastoma susceptibility protein"
/protein_id="AAAS3483.1"
/db_xref="GI:292421"

protein_bind

protein_bind

CDS

exon

intron

variation

/standard_name="Restriction Fragment Length Polymorphism"

```
exon      /note="BamHI RFLP"
5424..5550
intron    /number=2
5551..59445
          /number=2
satellite 14868..14914
          /standard_name="microsatellite RBi2 (AFM058xd6)"
exon      39446..39561
intron    /number=3
39562..41925
          /number=3
exon      41926..42045
intron    /number=4
42046..44667
          /number=4
satellite 43218..43257
          /standard_name="microsatellite RBi4"
exon      44668..44706
intron    /number=5
44707..45798
          /number=5
exon      45799..45866
intron    /number=6
45867..55852
          /number=6
exon      56853..56963
intron    /number=7
56964..59650
          /number=7
exon      59651..59793
intron    /number=8
59794..61729
          /number=8
exon      61730..61807
intron    /number=9
61808..64329
          /number=9
exon      64330..64439
intron    /number=10
64440..65363
          /number=10
exon      65364..65441
intron    /number=11
65442..70241
          /number=11
exon      70242..70329
intron    /number=12
70330..73752
          /number=12
exon      73753..73869
intron    /number=13
73870..76429
          /number=13
exon      76430..76486
intron    /number=14
76487..76888
          /number=14
exon      76889..76920
intron    /number=15
76921..77000
          /number=15
exon      77001..77077
intron    /number=16
77078..78082
          /number=16
exon      78083..78279
intron    /number=17
78280..149997
          /number=17
variation 99426
          /standard_name="Restriction Fragment Length Polymorphism"
          /note="XbaI RFLP"
satellite 102921..102959
```

```
repeat_region /standard_name="microsatellite RBi17"
123912..125501
          /standard_name="Variable Number Tandem Repeats (VNTR)"
          /rpt_type=tandem
exon      149998..150116
          /number=18
intron    150117..153207
          /number=18
exon      153208..153353
          /number=19
intron    153354..156692
          /number=19
exon      156693..156838
          /number=20
intron    156839..160729
          /number=20
satellite 156895..156976
          /standard_name="microsatellite RBi.20"
exon      160730..160834
          /number=21
intron    160835..161996
          /number=21
exon      161997..162110
          /number=22

Query Match      5.8%; Score 120.4; DB 9; Length 180388;
Best Local Similarity 89.0%; Pred. No. 1.4e-13;
Matches 130; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1943 CAAGTGTAAAGACTCTGCTGCACGACACACATATAAAAAAGAGAGAGAAATCAAAATAGAC 2002
      |||
Db 165647 CAAAATTGATAGACCGCTAGCAAGACTAATAAGAAAAAGAGAGAGAAATCAAAATAGAC 165706
      |||

QY 2003 ACAATATAAAAAATGATAAAGGGGATATCACACCGATCCACAGAAATCAAAATACCATC 2062
      |||
Db 165707 ACAATATAAAAAATGATAAAGGGGATATCACACCGATCCACAGAAATCAAAATACCATC 165766
      |||

QY 2063 AGAGAATACTACAAACACCTCTACGC 2088
      |||
Db 165767 AGAGATACTACAAACACCTCTACGC 165792
      |||

Search completed: September 9, 2005, 03:30:32
Job time : 6025.29 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 23:15:40 ; Search time 241.343 Seconds
(without alignments)
14156.413 Million cell updates/sec

Title: US-09-974-546C-83

Perfect score: 2088

Sequence: 1 gacctaataatatacgaggt.....tactacaacacctctacgc 2088

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2076	99.4	2087	3	US-09-097-199-83
2	1950	93.4	2505	3	US-09-097-199-83
3	1920.2	92.0	2506	4	US-09-949-016-1311
4	1820.2	92.0	6507	4	US-09-949-016-1311
5	572.2	27.4	601	4	US-09-949-016-15053
6	572.2	27.4	601	4	US-09-949-016-119365
7	344.2	16.5	601	4	US-09-949-016-119366
8	275.4	13.2	757	2	US-09-949-016-119364
9	275.4	13.2	757	3	US-08-692-787-3
10	118.8	5.7	146401	4	US-09-949-016-16151
11	117.2	5.6	162841	4	US-09-949-016-13733
12	115.6	5.5	573	3	US-09-385-982-420
13	115.2	5.5	430	3	US-09-397-787-248
14	115.2	5.5	475	4	US-09-621-976-12318
15	115.2	5.5	32042	3	US-09-245-281-44
16	115.2	5.5	32042	4	US-09-340-620-63
17	115.2	5.5	35837	4	US-09-949-016-15232
18	115.2	5.5	86439	4	US-09-949-016-11945
19	115.2	5.5	86440	4	US-09-949-016-16990
20	115.2	5.5	99370	4	US-09-949-016-12816
21	115.2	5.5	99370	4	US-09-949-016-17540
22	115.2	5.5	102520	4	US-09-949-016-17367
23	115.2	5.5	102526	4	US-09-949-016-12448
24	115.2	5.5	116592	4	US-09-818-512-3
25	115.2	5.5	124264	4	US-09-949-016-16396
26	115.2	5.5	168394	4	US-09-949-016-13002
27	115.2	5.5	175236	4	US-09-949-016-14353

C 28	115.2	5.5	206433	4	US-09-949-016-13527	Sequence 13527, A
C 29	115.2	5.5	254778	4	US-09-949-016-12417	Sequence 12417, A
C 30	115.2	5.5	312470	4	US-09-949-016-14043	Sequence 14043, A
C 31	115.2	5.5	336024	4	US-09-949-016-12373	Sequence 12373, A
C 32	115.2	5.5	363032	4	US-09-949-016-12415	Sequence 12415, A
C 33	115.2	5.5	363033	4	US-09-949-016-15754	Sequence 15754, A
C 34	113.6	5.4	439	3	US-09-397-787-269	Sequence 269, App
C 35	113.6	5.4	601	4	US-09-949-016-116730	Sequence 116730, A
C 36	113.6	5.4	4856	4	US-09-949-016-14956	Sequence 14956, A
C 37	113.6	5.4	69813	4	US-09-949-016-12455	Sequence 12455, A
C 38	113.6	5.4	69813	4	US-09-949-016-13905	Sequence 13905, A
C 39	113.6	5.4	69813	4	US-09-949-016-13906	Sequence 13906, A
C 40	113.6	5.4	69833	4	US-09-949-016-12861	Sequence 12861, A
C 41	113.6	5.4	80706	4	US-09-949-016-15347	Sequence 15347, A
C 42	113.6	5.4	87648	4	US-09-949-016-13655	Sequence 13655, A
C 43	113.6	5.4	95648	4	US-09-949-016-13139	Sequence 13139, A
C 44	113.6	5.4	106380	4	US-09-949-016-17553	Sequence 17553, A
C 45	113.6	5.4	107421	4	US-09-949-016-15532	Sequence 15532, A

ALIGNMENTS

RESULT 1
US-09-097-199-83
; Sequence 83, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2087 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 99..503
US-09-097-199-83

Query Match 99.4%; Score 2076; DB 3; Length 2087;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2087; Conservative 0; Mismatches 0; Indels 1; Gaps 1;			
QY	1	GACCTTAAATATATCAGGTGGCTTAATTCGATGATATTAATTAATTAACAAAATTTCTCTCTA	60
Db	1	GACCTTAAATATATCAGGTGGCTTAATTCGATGATATTAATTAATTAACAAAATTTCTCTCTA	60
QY	61	TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCAATGAGGGCTCTCTTAAGGAAC	120
Db	61	TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCAATGAGGGCTCTCTTAAGGAAC	120
QY	121	AGAAATATGAGGATATGCACAATTTATTTACATTTTACAGATCAGAAAATTTAGGACCA	180
Db	121	AGAAATATGAGGATATGCACAATTTATTTACATTTTACAGATCAGAAAATTTAGGACCA	180
QY	181	GATTAAAGTAACTTCCCAAGCTACCAAGGATTCAGCTCCAGAAAATGTCCTCTTACCAT	240
Db	181	GATTAAAGTAACTTCCCAAGCTACCAAGGATTCAGCTCCAGAAAATGTCCTCTTACCAT	240
QY	241	TCTGCTACAGGTATTTTCGAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT	300
Db	241	TCTGCTACAGGTATTTTCGAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT	300
QY	301	TCATTGATTTCCATAGAACAGTCACACATGCAATTTCTCACCCCTTGCGACACACT	360
Db	301	TCATTGATTTCCATAGAACAGTCACACATGCAATTTCTCACCCCTTGCGACACACT	360
QY	361	TGACCATGAAAGGTCCTCAATGAAATGTTCTCATTTATCTTCAGAAAGCCATATTATTC	420
Db	361	TGACCATGAAAGGTCCTCAATGAAATGTTCTCATTTATCTTCAGAAAGCCATATTATTC	420
QY	421	CATTGACTTTGCGAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTCTCTACTTATCCA	480
Db	421	CATTGACTTTGCGAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTCTCTACTTATCCA	480
QY	481	AAACTATACATCCAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTTCAGAAAA	540
Db	481	AAACTATACATCCAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTTCAGAAAA	540
QY	541	ATAAAAATGTTGAAAGGCAATTCCTGCTACCAATGACTGTTTAAAGCCGCAAGTAAAC	600
Db	541	ATAAAAATGTTGAAAGGCAATTCCTGCTACCAATGACTGTTTAAAGCCGCAAGTAAAC	600
QY	601	TGAACCATTTCCAACTTCAATTTACTTATGAAAGAAATTTGATGATAGGAGTTATTTTC	660
Db	601	TGAACCATTTCCAACTTCAATTTACTTATGAAAGAAATTTGATGATAGGAGTTATTTTC	660
QY	661	AATTCCTAAAATACAAACCCATGTTGATCTTTTCAATCTTGAACTCATAGATTATTTCT	720
Db	661	AATTCCTAAAATACAAACCCATGTTGATCTTTTCAATCTTGAACTCATAGATTATTTCT	720
QY	721	ATTATCTCAATTTAGTTTGTATTATCTAGTGGGCCATTAAAACTACCACATGTTT	780
Db	721	ATTATCTCAATTTAGTTTGTATTATCTAGTGGGCCATTAAAACTACCACATGTTT	780
QY	781	TCGTCTCTCCATTTAGTCAATTAACCTAAACGAGCAATTTAGTAAGCCATGTGCCAGAT	840
Db	781	TCGTCTCTCCATTTAGTCAATTAACCTAAACGAGCAATTTAGTAAGCCATGTGCCAGAT	840
QY	841	GCTCCGCTAGGCACAGAGGATAAAAACAATTTATAGTATACCACTAAATTTTCGCTT	900
Db	841	GCTCCGCTAGGCACAGAGGATAAAAACAATTTATAGTATACCACTAAATTTTCGCTT	900
QY	901	AGTAACTAGTGAATGTTCAAGTCATGCTGAGTCAAGAGTTGAGGAGACATTTACAAATGT	960
Db	901	AGTAACTAGTGAATGTTCAAGTCATGCTGAGTCAAGAGTTGAGGAGACATTTACAAATGT	960
QY	961	GTAATGGAAACCAAGGAAAGTGAACCTTTGGATAAGTGGGACTAGTGTATTATATTT	1020
Db	961	GTAATGGAAACCAAGGAAAGTGAACCTTTGGATAAGTGGGACTAGTGTATTATATTT	1020
QY	1021	TAAATGATTTCTGACTCTATCATTTGGCCCTCCAAACACAGATTTGTTTCTTTGTTT	1080

; Sequence 85, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltre, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2505 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:

; NAME/KEY: CDS
; LOCATION: 99..503
; US-09-199-85
Query Match 93.4%; Score 1950; DB 3; Length 2505;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY	1	GACCTTAATATATCGAGGTGGCTAATTGATGTATATAATTTACAAAATTTATTTCTTA	60
Db	1	GACCTTAATATATCGAGGTGGCTAATTGATGTATATAATTTACAAAATTTATTTCTTA	60
QY	61	TTGCTACAGAGTACAAATTCAAATTTACAGPAGGCCACCAATAGGGCTTTTAAAGAAC	120
Db	61	TTGCTACAGAGTACAAATTCAAATTTACAGPAGGCCACCAATAGGGCTTTTAAAGAAC	120
QY	121	AGAAATATGAGGATATGCACAAATTTATTTACATTTTACAGATCAGAAAATTTAGGCACA	180
Db	121	AGAAATATGAGGATATGCACAAATTTATTTACATTTTACAGATCAGAAAATTTAGGCACA	180
QY	181	GATTAAGTAACTTCCCAAGGCTACCAAGCATTTAGCTCCAGAAAATTTAGTACCAT	240
Db	181	GATTAAGTAACTTCCCAAGGCTACCAAGCATTTAGCTCCAGAAAATTTAGTACCAT	240
QY	241	TTGCTACAGAGTATTTTTCGAAAAAGAAAAGTAAAAAGAGTAAAAAGGCAACAGAGT	300
Db	241	TTGCTACAGAGTATTTTTCGAAAAAGAAAAGTAAAAAGAGTAAAAAGGCAACAGAGT	300
QY	301	TCATTGATTTTCCATAGAACAGTACACCATCGAATTTCTCACACCTTTGCAGACACT	360

Db	301	TCATTGATTTTCCATAGAACAGTACACCATCGAATTTCTCACACCTTTGCAGACACT	360
QY	361	TGACCATGAAAGGTTCTCTCAATGAAATGTTCTCTCAATTTCTTTCAGAAAGCATATTATTC	420
Db	361	TGACCATGAAAGGTTCTCTCAATGAAATGTTCTCTCAATTTCTTTCAGAAAGCATATTATTC	420
QY	421	CATTGACTTTGCAGTTAACTCAGACCCCTAGGTCTGGAATGCTGTCTCTCTACTTATTC	480
Db	421	CATTGACTTTGCAGTTAACTCAGACCCCTAGGTCTGGAATGCTGTCTCTCTACTTATTC	480
QY	481	AAACTATATACATCACAGATCATATAAACTCTCAGCCCTCTGCAAAAGCCTTTTCCAGAAA	540
Db	481	AAACTATATACATCACAGATCATATAAACTCTCAGCCCTCTGCAAAAGCCTTTTCCAGAAA	540
QY	541	ATAAAAATGTTGAAAAGGCAATTTCTGCTACCAATGACTGTTTAAAGCCAGCCAACTAAC	600
Db	541	ATAAAAATGTTGAAAAGGCAATTTCTGCTACCAATGACTGTTTAAAGCCAGCCAACTAAC	600
QY	601	TGAACCATTTCCAACTTCAATTTTACATGAAAAGAAATTTGATGATGTAGGAGGTTATTC	660
Db	601	TGAACCATTTCCAACTTCAATTTTACATGAAAAGAAATTTGATGATGTAGGAGGTTATTC	660
QY	661	AATTTCTAAAATACAAACCCATGTTGATCTTCTCAATCTTGAATCTCATAGATTATTTATCT	720
Db	661	AATTTCTAAAATACAAACCCATGTTGATCTTCTCAATCTTGAATCTCATAGATTATTTATCT	720
QY	721	ATTATCTCAATTTAGTTTGTATTTATCTAGTGGGCCATTTAAAACTACCAATGTTT	780
Db	721	ATTATCTCAATTTAGTTTGTATTTATCTAGTGGGCCATTTAAAACTACCAATGTTT	780
QY	781	TCTGTCTCTCAATTTAGTCAATTAACCTAACAGAGCAATTTAGTAAAGCCATGTCAGAT	840
Db	781	TCTGTCTCTCAATTTAGTCAATTAACCTAACAGAGCAATTTAGTAAAGCCATGTCAGAT	840
QY	841	GCTCCGCTAGGCAACAGAGGATAAAAACAATTTATAGTATACCACCTAAATTTTCGCTT	900
Db	841	GCTCCGCTAGGCAACAGAGGATAAAAACAATTTATAGTATACCACCTAAATTTTCGCTT	900
QY	901	AGTAACTAGTGAATTTTCAAGTCTAGTCAAGAGTTGAGGAGACATTTACAAATGT	960
Db	901	AGTAACTAGTGAATTTTCAAGTCTAGTCAAGAGTTGAGGAGACATTTACAAATGT	960
QY	961	GTAATGGAACCAAGGAAAGTGAATTTGGAATAGTGGGACTAGTGTATTTATATAT	1020
Db	961	GTAATGGAACCAAGGAAAGTGAATTTGGAATAGTGGGACTAGTGTATTTATATAT	1020
QY	1021	TAATTTGATTTCTGACTCTATCAATTTGSCCTCCAAACACAGATTTGTTTCTTTGTTT	1080
Db	1021	TAATTTGATTTCTGACTCTATCAATTTGSCCTCCAAACACAGATTTGTTTCTTTGTTT	1080
QY	1081	GTTTTCTTCCATATGGGATCTTCTGTGCCAGACAGTGCCTGACACATAGAAAACAATC	1140
Db	1081	GTTTTCTTCCATATGGGATCTTCTGTGCCAGACAGTGCCTGACACATAGAAAACAATC	1140
QY	1141	AATATTTGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1200
Db	1141	AATATTTGCTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1200
QY	1201	AGAACTCCAGAGTAAAGTACAGGSCCTCTGCATTTATATGCGCTTAAATTAAGATTAT	1260
Db	1201	AGAACTCCAGAGTAAAGTACAGGSCCTCTGCATTTATATGCGCTTAAATTAAGATTAT	1260
QY	1261	GTGAGAAAATGTTTAAAGACACTTAGTAGAGTGAATTTTGAATATATAGTAAACACTT	1320
Db	1261	GTGAGAAAATGTTTAAAGACACTTAGTAGAGTGAATTTTGAATATATAGTAAACACTT	1320
QY	1321	TGGTGGTGTCTTAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1380
Db	1321	TGGTGGTGTCTTAAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1380
QY	1381	CATAAATATTTTAAAGGAAAATCAGATCTCCAGGCTTTCAATGTTTGTTCATTTT	1440
Db	1381	CATAAATATTTTAAAGGAAAATCAGATCTCCAGGCTTTCAATGTTTGTTCATTTT	1440

1441 CATATATTTTACCATCTGCTGAAGGAGGTATATCAATCAAGGGTAAAGAAAGATGGGAGGA 1500
 1441 CATATATTTTACCATCTGCTGAAGGAGGTATATCAATCAAGGGTAAAGAAAGATGGGAGGA 1500
 1501 AAACCTCAGTAAGAAATATATAGTCTGTTTGGCAAGGTAGAAAAGATTTCTCATCACTCAA 1560
 1501 AAACCTCAGTAAGAAATATATAGTCTGTTTGGCAAGGTAGAAAAGATTTCTCATCACTCAA 1560
 1561 CCTTATGAGCAGGAAGGAGGCTGTTTGGCAAGGTAGAAAAGATTTCTCATCACTCAA 1620
 1561 CCTTATGAGCAGGAAGGAGGCTGTTTGGCAAGGTAGAAAAGATTTCTCATCACTCAA 1620
 1621 TTTAGACACTTCCCTGCATTAATGCTGACCAAAATATGTTTGGCAAGGTAGAAAAGATTTCTCATCACTCAA 1680
 1621 TTTAGACACTTCCCTGCATTAATGCTGACCAAAATATGTTTGGCAAGGTAGAAAAGATTTCTCATCACTCAA 1679
 1681 CTCCTCAACAGCAGCAATTCAGGAGTTAAATATTTTTCATCAAAATTTTTCCTTTA 1740
 1680 CTCCTCAACAGCAGCAATTCAGGAGTTAAATATTTTTCATCAAAATTTTTCCTTTA 1739
 1741 AGCTAGAGATGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTTAAGAGAGAT 1800
 1740 AGCTAGAGATGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTTAAGAGAGAT 1799
 1801 TTCTCCCTGTTTAAAGCAGCAAGCAAAATTTAGGCAATTTCACTCTCAAACTTCACTAAATGA 1860
 1800 TTCTCCCTGTTTAAAGCAGCAAGCAAAATTTAGGCAATTTCACTCTCAAACTTCACTAAATGA 1859
 1861 TCACATTTCTTCCAAAGGAACCTCTAGAGACCAAAATGCCCCGAGTTAAAGACATCAAAA 1920
 1860 TCACATTTCTTCCAAAGGAACCTCTAGAGACCAAAATGCCCCGAGTTAAAGACATCAAAA 1919
 1921 CTAAACCTCTGAAGAACTTCCCAAGTGAAGACTCTGCTGCAAGCAACATCAAAA 1978
 1920 CTAAACCTCTGAAGAACTTCCCAAGTGAAGACTCTGCTGCAAGCAACATCAAAA 1977

RESULT 3

US-09-949-016-3311
 ; Sequence 3311, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949, 016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 3311
 ; LENGTH: 2506
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-3311

Query Match 92.0%; Score 1920.2; DB 4; Length 2506;
 Best Local Similarity 99.1%; Pred. No. 0;
 Matches 1961; Conservative 1; Mismatches 14; Indels 3; Gaps 3;
 1 GACCTTAAATATATCGAGGTGGCTAATGATGATATATATATTAATTTACAAAATTTCTTCTTA 60
 1 GACCTTAAATATATCGAGGTGGCTAATGATGATATATATATTAATTTACAAAATTTCTTCTTA 60
 61 TTGCTACAGAGCTACAAATTCAAATTTACGTAGGCAACCATGAGGGCTTCTTAAGGAACC 120

Db QY 118 -TGCTACAGAGCTACAAATTCAAATTTACAGTA -GTAAATATAGAGGCTTCTTTAAGGAACC 118
 QY 121 AGAAATATGAGGATATGACAAATATATTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
 Db 119 AGAAATATGAGGATATGACAAATATATTTACATTTTACAGATCAGAAAATTTGAGGCACA 178
 QY 181 GATTAAAGTAACTTTCCCAAGGCTTACAGGCAATTTCTAGCTCCAGAAAATTTGCTCTTACCAT 240
 Db 179 GATTAAAGTAACTTTCCCAAGGCTTACAGGCAATTTCTAGCTCCAGAAAATTTGCTCTTACCAT 238
 QY 241 TCTGCTACAAAGTATTTTGGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 300
 Db 239 TCTGCTACAAAGTATTTTGGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 298
 QY 301 TCATTTGATTTATTCATAGAACAGTCAACCATGCAATTTCTCACCCCTTGCAGACACACT 360
 Db 299 TCATTTGATTTATTCATAGAACAGTCAACCATGCAATTTCTCACCCCTTGCAGACACACT 358
 QY 361 TGACCATGAAAGGTTTCCCTCAATGAAATGTTCTCATTTATCTTTCAGAAAGCATATTTATTTCA 420
 Db 359 TGACCATGAAAGGTTTCCCTCAATGAAATGTTCTCATTTATCTTTCAGAAAGCATATTTATTTCA 418
 QY 421 CATTTGACTTTGAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTTCTCTACTTATCCA 480
 Db 419 CATTTGACTTTGAGTTAACTCAGACCCCTAGGCTGGAATGCTGCTTCTCTACTTATCCA 478
 QY 481 AAACATATACATCCACAGATCATATAAATCTCTCAGCCCTGCTGCAAGGCTTTCCAGAAA 540
 Db 479 AAACATATACATCCACAGATCATATAAATCTCTCAGCCCTGCTGCAAGGCTTTCCAGAAA 538
 QY 541 ATAAAAATGTTTGAAGGCAATTTCTGCTACCAATGCACTGTTTAAAGCCAGCCAAAGTAAC 600
 Db 539 ATAAAAATGTTTGAAGGCAATTTCTGCTACCAATGCACTGTTTAAAGCCAGCCAAAGTAAC 598
 QY 601 TGAAACATTTCAAATTTCAATTTTACATTTGATGAAAAGAAATTTGATGATGATGAGGTTTATTC 660
 Db 599 TGAAACATTTCAAATTTCAATTTTACATTTGATGAAAAGAAATTTGATGATGATGAGGTTTATTC 658
 QY 661 AATTTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACCTCATAGATTTATTTCT 720
 Db 659 AATTTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACCTCATAGATTTATTTCT 718
 QY 721 ATTTCTCAATTTAGTTTGTATTTATTTCTAGTGGGCCATTTAAAAAATACCAATGTTT 780
 Db 719 ATTTCTCAATTTAGTTTGTATTTATTTCTAGTGGGCCATTTAAAAAATACCAATGTTT 778
 QY 781 TCTGCTCTCTCAATTTAGTCTAATTAACCTAAACCTAAACGAGCAATTTAGTAAAGCATGTCAGAT 840
 Db 779 TCTGCTCTCTCAATTTAGTCTAATTAACCTAAACCTAAACGAGCAATTTAGTAAAGCATGTCAGAT 838
 QY 841 GCTCCGCTAGGCACAGAGGGGATAAAAAATACCTTATAGTATACCACTAAATTTTTCGCTT 900
 Db 839 GCTCCGCTAGGCACAGAGGGGATAAAAAATACCTTATAGTATACCACTAAATTTTTCGCTT 898
 QY 901 AGTAACTAGTGAATTTTCAAGTCTAGTCTGAGTCAAGAGTTTGAAGGAGACATTTACAATGT 960
 Db 899 AGTAACTAGTGAATTTTCAAGTCTAGTCTGAGTCAAGAGTTTGAAGGAGACATTTACAATGT 958
 QY 961 GTAATGGAACCAAGGAAAGTGAATTTTGGATAAGTGGGAGCTAGTGTATTTATATATT 1020
 Db 959 GTAATGGAACCAAGGAAAGTGAATTTTGGATAAGTGGGAGCTAGTGTATTTATATATT 1018
 QY 1021 TAAATGATTTTCTGACTCTATCATTTGGCTCCCAAAACACAGATTTGTTTTCTTTGGTTTT 1080
 Db 1019 TAAATGATTTTCTGACTCTATCATTTGGCTCCCAAAACACAGATTTGTTTTCTTTGGTTTT 1078
 QY 1081 GTTTTCTTCTCAATGGATCTTTCTGTGCCAGCAGCATGTCCTGACACATAGAAAAACAATC 1140
 Db 1079 GTTTTCTTCTCAATGGATCTTTCTGTGCCAGCAGCATGTCCTGACACATAGAAAAACAATC 1138
 QY 1141 -AATATTTGCTGAATTAATGATTTAAAAATCAGAGAACTTTTCCCATTTCTGTTTGGATCTA 1199
 Db 1139 AATATTTGCTGAATTAATGATTTAAAAATCAGAGAACTTTTCCCATTTCTGTTTGGATCTA 1198

QY 961 GTAATGGAACCAAGGAAAGTGAACCTTTGGATAAGTGGGACTAGTGATTTATATATT 1020
Db |||||
QY 2959 GTAATGGAACCAAGGAAAGTGAACCTTTGGATAAGTGGGACTAGTGATTTATATATT 3018
Db |||||
QY 1021 TAAATGATTTCTGACTCTATCATTTGGCCCTCCAAAACACAGATGTGTTTTCTTGGTTTT 1080
Db |||||
QY 3019 TAAATGATTTCTGACTCTATCATTTGGCCCTCCAAAACACAGATGTGTTTTCTTGGTTTT 3078
Db |||||
QY 1081 GTTTTCTTCACTATGGGATCTTCTGGCCCGACACAGTGCCTGCACATAGAAAACATC 1140
Db |||||
QY 3079 GTTTTCTTCACTATGGGATCTTCTGGCCCGACACAGTGCCTGCACATAGAAAACATC 3138
Db |||||
QY 1141 -AATATTTCTGAAATAAATGATTAATAAATCAGAGAACTTTCCCAATTCGTGTTGGATCTA 1199
Db |||||
QY 3139 AATATTTCTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCAATTCGTGTTGGATCTA 3198
Db |||||
QY 1200 TAGAACATCCAGAGTAAGTATGAGGCTCTGCAATTTATATGCGCTTAAATAAGATTA 1259
Db |||||
QY 3199 TAGAACATCCAGAGTAAGTATGAGGCTCTGCAATTTATATGCGCTTAAATAAGATTA 3258
Db |||||
QY 1260 TGTGAGAAAGTTTAAAGACACTTAGTAGAGGATTTTGAAATATAGTAAACACTTGGAA 1319
Db |||||
QY 3259 TGTGAGAAAGTTTAAAGACACTTAGTAGAGGATTTTGAAATATAGTAAACACTTGGAA 3318
Db |||||
QY 1320 ATGGTGGTCTTTAAAGAGATTAATATAGATTAATATGAAATCTCCATCTCAAAATAAT 1379
Db |||||
QY 3319 ATGGTGGTCTTTAAAGAGATTAATATAGATTAATATGAAATCTCCATCTCAAAATAAT 3378
Db |||||
QY 1380 GATATACTATTAAGGAAATATCATCTCCAGGCTTTCAATGTTTGTTCATTAATTTT 1439
Db |||||
QY 3379 GCATAAATCTATTAAGGAAATATCATCTCCAGGCTTTCAATGTTTGTTCATTAATTTT 3438
Db |||||
QY 1440 TCATATATTTTACCATCTGCTGAAGGAGTATATCAAGGCTTAAAGAAAGATGGGAGG 1499
Db |||||
QY 3439 TCATATATTTTACCATCTGCTGAAGGAGTATATCAAGGCTTAAAGAAAGATGGGAGG 3498
Db |||||
QY 1500 AAAAATCAGTAAGAAATTAATATATGATCTGTTTGCAGGAGTAAAGAAAGATTTCTCATCTCA 1559
Db |||||
QY 3499 AAAAATCAGTAAGAAATTAATATATGATCTGTTTGCAGGAGTAAAGAAAGATTTCTCATCTCA 3558
Db |||||
QY 1560 ACCTTATGAGGAGGAGGAGGAGGCTGTTTGAAGACCAATTTACTTTAGCAGAACACATATA 1619
Db |||||
QY 3559 ACCTTATGAGGAGGAGGAGGAGGCTGTTTGAAGACCAATTTACTTTAGCAGAACACATATA 3618
Db |||||
QY 1620 TTTTGAACATTTCCCTGCAATTAACATGCAACAATAATGTTTGCAGGAGTAAAGAAAGATTTCTCATCTCA 1679
Db |||||
QY 3619 TTTTGAACATTTCCCTGCAATTAACATGCAACAATAATGTTTGCAGGAGTAAAGAAAGATTTCTCATCTCA 3678
Db |||||
QY 1680 CCTCCAAACAGACACATTTAGGAGTAAATATTTTTCATCAAAATTTGGATTTTTCCTTT 1739
Db |||||
QY 3679 CCTCCAAACAGACACATTTAGGAGTAAATATTTTTCATCAAAATTTGGATTTTTCCTTT 3738
Db |||||
QY 1740 AACGCTAGAGATGCTCAAAATCTTCTGAGGAGTCAATGCGCTTCAAGCTAAGAGAGATA 1799
Db |||||
QY 3739 AACGCTAGAGATGCTCAAAATCTTCTGAGGAGTCAATGCGCTTCAAGCTAAGAGAGATA 3798
Db |||||
QY 1800 TTTTCTCCCTGTTTAAAGCAGCAAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATG 1859
Db |||||
QY 3799 TTTTCTCCCTGTTTAAAGCAGCAAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATG 3858
Db |||||
QY 1860 ATCAATCTTTTCCAAAAGGAACTCTAGAGAACCAATGCCCCGAGTTTAAAGAACATCAAA 1919
Db |||||
QY 3859 ATCAATCTTTTCCAAAAGGAACTCTAGAGAACCAATGCCCCGAGTTTAAAGAACATCAAA 3918
Db |||||
QY 1920 ACTAACCATCTGAGAACTTCCGAGGTAGACTCTGCGCTGCGAGGACACATATAA 1978
Db |||||
QY 3919 ACTAACCATCTGAGAACTTCCGAGGTAGACTCTGCGCTGCGAGGACATTAACCGAGA 3977

RESULT 5
US-09-949-016-119365/c
; Sequence 119365, Application US/09949016

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119365
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-119365

Query Match 27.4%; Score 572.2; DB 4; Length 601;
Best Local Similarity 98.8%; Pred. No. 1.6e-132;
Matches 596; Conservative 1; Mismatches 4; Indels 2; Gaps 2;

QY 2 ACCTTAAATATATCGAGTGGCTAATTGATGTAATAAATTTACAAAATTTCTTCTAT 61
Db |||||
QY 601 ACCTTAAATATATCGAGTGGCTAATTGATGTAATAAATTTACAAAATTTCTTCTA- 543
Db |||||
QY 62 TGCTACAGAGCTACAATTTCAATTTTACAGTGGCCACCATGAGGGCTTCTTTAAGGAACA 121
Db |||||
QY 542 TGCTACAGAGCTACAATTTCAATTTTACAGT-ATAATATGAGGGCTTCTTTAAGGAACA 484
Db |||||
QY 122 GAAATATGAGATATGCAATATTTTCAATTTTACAGATCGAATAATTTGAGGCAAG 181
Db |||||
QY 483 GAAATATGAGATATGCAATATTTTCAATTTTACAGATCGAATAATTTGAGGCAAG 424
Db |||||
QY 182 ATTAAGTAACTTTCCAGGCTACAGGCTTCTAGTCTCAGAAATGTCGTCTTACCAAT 241
Db |||||
QY 423 ATTAAGTAACTTTCCAGGCTACAGGCTTCTAGTCTCAGAAATGTCGTCTTACCAAT 364
Db |||||
QY 242 CTGCTACAAGGTATTTTCAAAAGGAAAGTAAAGAAAGTCAAAAGGCAACAGAGTT 301
Db |||||
QY 363 CTGCTACAAGGTATTTTCAAAAGGAAAGTAAAGAAAGTCAAAAGGCAACAGAGTT 304
Db |||||
QY 302 CATTGATTTTCCATAGAACAGTCAACCATGCAATTTCTCACCCCTTGCAGACACACTT 361
Db |||||
QY 303 CATTGATTTTCCATAGAACAGTCAACCATGCAATTTCTCACCCCTTGCAGACACACTT 244
Db |||||
QY 362 GACCATGAAAGGTTCTCAATGAAATGTTCTCTATTATCTTTCAGAGCCATATTATTCAC 421
Db |||||
QY 243 GACCATGAAAGGTTCTCAATGAAATGTTCTCTATTATCTTTCAGAGCCATATTATTCAC 184
Db |||||
QY 422 ATTGACTTTGAGTTAACTCAGACCCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAA 481
Db |||||
QY 183 ATTGACTTTGAGTTAACTCAGACCCCTAGGTCTGGAATGCTGTCTTCTCTACTTATCCAA 124
Db |||||
QY 482 AACTATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTTCAGAAAAA 541
Db |||||
QY 123 AACTATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTTCAGAAAAA 64
Db |||||
QY 542 TAAAAATGGTTGAAAAAGGCAATTTCTGCTACCAATGACTGTTTAAAGCCCGAGCCAGTAAC 601
Db |||||
QY 63 TAAAAATGGTTGAAAAAGGCAATTTCTGCTACCAATGACTGTTTAAAGCCCGAGCCAGTAAC 4
Db |||||
QY 602 GAA 604
Db |||||
QY 3 GAA 1

RESULT 6
US-09-949-016-119366/c

; Sequence 119366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119366
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119366

Query Match 27.4%; Score 572.2; DB 4; Length 601;
Best Local Similarity 98.8%; Pred. No. 1.6e-132;
Matches 596; Conservative 1; Mismatches 4; Indels 2; Gaps 2;
QY 14 TCAGGTGGCTAATTCATGTATATAATTTACAAAATTTCTTCTATTGCTACAGAGCT 73
Db 601 TCAGGTGGCTAATTCATGTATATAATTTACAAAATTTCTTCTA-TGCTACAGAGCT 543
QY 74 ACAATTCATTTACAGTAGGCCACCATGAGGGCTTCTTAAGAACAGAAATATGAGGA 133
Db 542 ACAATTCATTTACAGTA-GTAAATATGAGGGCTTCTTAAGAACAGAAATATGAGGA 484
QY 134 TATGCAATATTTATTTACATTTTACAGATCAGAAATTTGAGCAGAGTAAGTAACCT 193
Db 483 TATGCAATATTTATTTACATTTTACAGATCAGAAATTTGAGCAGAGTAAGTAACCT 424
QY 194 CCCAAGGCTACCAAGGCAATTTAGCTCCAGAAATCTGTCTTTACCAATTCGTCTACAAGGT 253
Db 423 CCCAAGGCTACCAAGGCAATTTAGCTCCAGAAATCTGTCTTTACCAATTCGTCTACAAGGT 364
QY 254 ATTTGAAAAAAGAAAAAGTAAAGAGTCAAAAGGCAACAGAGTTCATTTGATTTTC 313
Db 363 ATTTGAAAAAAGAAAAAGTAAAGAGTCAAAAGGCAACAGAGTTCATTTGATTTTC 304
QY 314 CATAGAACAGTCACACCATGCAATTTCTCACACCTTGACAGACACACTTGACCATGAAAG 373
Db 303 CAYAGAACAGTCACACCATGCAATTTCTCACACCTTGACAGACACACTTGACCATGAAAG 244
QY 374 TTCTCTCAATGAAATGTTCTCTCATTTCTTTAGAGCCATATTTATTCACATTTGACTTTGCA 433
Db 243 TTCTCTCAATGAAATGTTCTCTCATTTCTTTAGAGCCATATTTATTCACATTTGACTTTGCA 184
QY 434 GTTAATCAGACCTTAGGTCTGGAATGCTGTCTCTACTTATCCAAAATATACATCC 493
Db 183 GTTAATCAGACCTTAGGTCTGGAATGCTGTCTCTACTTATCCAAAATATACATCC 124
QY 494 ACAGATCATATAAATCTCTCAGCCCTCTCAAGGCTTTCCAGAAAAATAAAAAATGTTG 553
Db 123 ACAGATCATATAAATCTCTCAGCCCTCTCAAGGCTTTCCAGAAAAATAAAAAATGTTG 64
QY 554 AAAAGGCAATTTCTGTACCAATGACTGTTTAAAGCCAGCCCAAGTAAGTCAATTCCTCAA 613
Db 63 AAAAGGCAATTTCTGTACCAATGACTGTTTAAAGCCAGCCCAAGTAAGTCAATTCCTCAA 4
QY 614 CTT 616
Db 3 CTT 1

RESULT 7

US-09-949-016-119364/c
; Sequence 119364, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119364
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119364

Query Match 16.5%; Score 344.2; DB 4; Length 601;
Best Local Similarity 98.1%; Pred. No. 7.7e-76;
Matches 368; Conservative 1; Mismatches 4; Indels 2; Gaps 2;
QY 1 GACCTTAATAATATATCGAGTGGCTAATTTGATGTATAATAATTTACAAAATTTCTTCTA 60
Db 373 GACCTTAATAATATATCGAGTGGCTAATTTGATGTATAATAATTTACAAAATTTCTTCTA 314
QY 61 TTGCTACAGAGCTACAATTTCAATTTACAGTAGGCCACCATGAGGGCTTCTTAAGGAACC 120
Db 313 -TGCTACAGAGCTCAATTTCAATTTACAGTA-GTAAATATGAGGGCTTCTTAAGGAACC 256
QY 121 AGAAATATGAGGATATGACATATATTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
Db 255 AGAAATATGAGGATATGACATATATTTACATTTTACAGATCAGAAAATTTGAGGCACA 196
QY 181 GATTAAAGTAATTTCCCAAGGCTACCAAGGCTTCTAGCTCCAGAAATCTGTCTTTACCAT 240
Db 195 GATTAAAGTAATTTCCCAAGGCTTACCAAGGCTTCTAGCTCCAGAAATCTGTCTTTACCAT 136
QY 241 TCTGCTACAAGGTATTTCCGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 300
Db 135 TCTGCTACAAGGTATTTCCGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 76
QY 301 TCATTGATTTTCCATAGAACAGTCACACCATGCAATTTCTCACACCTTTGACAGACACT 360
Db 75 TCATTGATTTTCCATAGAACAGTCACACCATGCAATTTCTCACACCTTTGACAGACACT 16
QY 361 TGACCATGAAAGGTT 375
Db 15 TGACCATGAAAGGTT 1

RESULT 8

US-08-692-787-3
; Sequence 3, Application US/08692787
; Patent No. 5882864
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE
; DISEASE
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433

/ CITY: Houston
/ STATE: Texas
/ COUNTRY: United States of America
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/692,787
/ FILING DATE: Concurrently Herewith
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Corder, Timothy S.
/ REGISTRATION NUMBER: 38,414
/ REFERENCE/DOCKET NUMBER: UROC:012
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 757 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-08-692-787-3

Query Match 13.2%; Score 275.4; DB 2; Length 757;
Best Local Similarity 96.2%; Pred. No. 1.1e-58;
Matches 282; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1686 ACAAGCACATTCAGGAGTTAAATATTTTCATCAAAACATTTGGATTTTCTTAAACGCT 1745
Db 1 ACAAGCACATTCAGGAGTTAAATATTTTCATCAAAACATTTGGATTTTCTTAAACGCT 60
QY 1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAGAGATTTCTC 1805
Db 61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAGAGATTTCTC 120
QY 1806 CTTGTATTAAGCAGCAGCAAAATAGCCATTTCATCTCAAACTTCACTAATGATCACA 1865
Db 121 CTTGTATTAAGCAGCAGCAAAATAGCCATTTCATCTCAAACTTCACTAATGATCACA 180
QY 1866 TTCTTCCAAAAGGAACCTCTAGAAGCACCATAATGCCCGAGTTAAGAACATCAAAACTAAC 1925
Db 181 TTCTTCCAAAAGGAACCTCTAGAAGCACCATAATGCCCGAGTTAAGAACATCAAAACTAAC 240
QY 1926 CATCTGAAGAAACTTCCCAAGTGAAGACTCTGCCTGCACGACACACATATAA 1978
Db 241 CATCTGAAGAAACTTCCCAAGTGAAGACTCTGCCTGCACGACATTAACCGAGA 293

RESULT 9
US-09-097-199-3
/ Sequence 3, Application US/09097199
/ Patent No. 6218529
/ GENERAL INFORMATION:
/ APPLICANT: An, Gang
/ APPLICANT: O'Hara, S. Mark
/ APPLICANT: Ralph, David
/ APPLICANT: Veltre, Robert
/ TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
/ PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
/ NUMBER OF SEQUENCES: 87
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Arnold, White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/097,199
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/692,787
/ FILING DATE: 31-JUL-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 757 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-09-097-199-3

Query Match 13.2%; Score 275.4; DB 3; Length 757;
Best Local Similarity 96.2%; Pred. No. 1.1e-58;
Matches 282; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1686 ACAAGCACATTCAGGAGTTAAATATTTTCATCAAAACATTTGGATTTTCTTAAACGCT 1745
Db 1 ACAAGCACATTCAGGAGTTAAATATTTTCATCAAAACATTTGGATTTTCTTAAACGCT 60
QY 1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAGAGATTTCTC 1805
Db 61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTAAGAGAGATTTCTC 120
QY 1806 CTTGTATTAAGCAGCAGCAAAATAGCCATTTCATCTCAAACTTCACTAATGATCACA 1865
Db 121 CTTGTATTAAGCAGCAGCAAAATAGCCATTTCATCTCAAACTTCACTAATGATCACA 180
QY 1866 TTCTTCCAAAAGGAACCTCTAGAAGCACCATAATGCCCGAGTTAAGAACATCAAAACTAAC 1925
Db 181 TTCTTCCAAAAGGAACCTCTAGAAGCACCATAATGCCCGAGTTAAGAACATCAAAACTAAC 240
QY 1926 CATCTGAAGAAACTTCCCAAGTGAAGACTCTGCCTGCACGACACACATATAA 1978
Db 241 CATCTGAAGAAACTTCCCAAGTGAAGACTCTGCCTGCATTAACCGAGA 293

RESULT 10
US-09-949-016-16151
/ Sequence 16151, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CLO01307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 16151
/ LENGTH: 146401
/ TYPE: DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2005, 03:30:51 ; Search time 1108.99 Seconds
(without alignments)
12364.714 Million cell updates/sec

Title: US-09-974-546C-83

Perfect score: 2088

Sequence: 1 gacctaataatatacgaggt.....tactacaacacctctacgc 2088

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq*
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq*
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq*
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq*
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq*
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq*
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq*
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq*
9:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq*
10:	/cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq*
11:	/cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq*
12:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq*
13:	/cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq*
14:	/cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq*
15:	/cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq*
16:	/cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq*
17:	/cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq*
18:	/cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq*
19:	/cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq*
20:	/cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq*
21:	/cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq*
22:	/cgn2_6/ptodata/2/pubpna/US10J_PUBCOMB.seq*
23:	/cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq*
24:	/cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq*
25:	/cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq*
26:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2076	99.4	2087	10	US-09-974-546-83
2	1950	93.4	2505	10	US-09-974-546-85
3	275.4	13.2	757	10	US-09-974-546-3
C 4	115.6	5.5	573	10	US-09-871-161-420
C 5	115.2	5.5	430	9	US-09-876-889-248
6	115.2	5.5	4180	11	US-09-984-429-334
7	115.2	5.5	4906	16	US-10-082-828A-143

8	115.2	5.5	4906	24	US-11-057-807-143	Sequence 143, Appl
9	115.2	5.5	6007	21	US-10-278-698-245	Sequence 245, Appl
10	115.2	5.5	6007	21	US-10-278-698-759	Sequence 759, Appl
11	115.2	5.5	6009	11	US-09-984-429-339	Sequence 339, Appl
12	115.2	5.5	6052	11	US-09-984-429-425	Sequence 425, Appl
13	115.2	5.5	6052	11	US-09-984-429-447	Sequence 447, Appl
14	115.2	5.5	6075	21	US-10-800-322-32	Sequence 32, Appl
15	115.2	5.5	6539	14	US-10-025-201-1	Sequence 1, Appl
16	115.2	5.5	7626	13	US-10-001-835-82	Sequence 82, Appl
17	115.2	5.5	11661	11	US-09-984-429-345	Sequence 345, Appl
C 18	115.2	5.5	30515	9	US-09-764-847-1208	Sequence 1208, Ap
C 19	115.2	5.5	30515	14	US-10-092-154-1208	Sequence 63, Appl
C 20	115.2	5.5	32042	9	US-09-728-721-63	Sequence 44, Appl
C 21	115.2	5.5	32042	13	US-10-118-984-44	Sequence 63, Appl
C 22	115.2	5.5	32042	15	US-10-295-981-63	Sequence 63, Appl
C 23	115.2	5.5	32042	21	US-10-843-188-63	Sequence 15, Appl
C 24	115.2	5.5	34001	15	US-10-006-883A-15	Sequence 1, Appl
C 25	115.2	5.5	68571	17	US-10-401-194-1	Sequence 6771, Ap
C 26	115.2	5.5	86500	20	US-10-719-993-6771	Sequence 3, Appl
C 27	115.2	5.5	116592	9	US-09-818-512-3	Sequence 3, Appl
C 28	115.2	5.5	116592	15	US-10-354-065-3	Sequence 3, Appl
C 29	115.2	5.5	116592	24	US-11-074-646-3	Sequence 3, Appl
C 30	115.2	5.5	147300	22	US-10-723-681-3	Sequence 274, App
C 31	115.2	5.5	227931	17	US-10-085-117-274	Sequence 274, App
C 32	115.2	5.5	227931	17	US-10-085-117-274	Sequence 274, App
C 33	115.2	5.5	383432	22	US-10-737-082-34	Sequence 34, Appl
C 34	115.2	5.5	383432	22	US-10-765-790-34	Sequence 34, Appl
C 35	115.2	5.5	402850	10	US-09-844-653-5	Sequence 5, Appl
C 36	115.2	5.5	561515	19	US-10-741-601-5682	Sequence 5682, Ap
C 37	115.2	5.5	561515	21	US-10-741-600-17730	Sequence 17730, A
C 38	115.2	5.5	606398	20	US-10-719-993-6782	Sequence 6782, Ap
C 39	115.2	5.5	1503841	9	US-09-795-668-1	Sequence 1, Appl
C 40	115.2	5.5	1503841	9	US-09-795-668-1	Sequence 1, Appl
C 41	115.2	5.5	1503841	9	US-09-795-668-1	Sequence 1, Appl
C 42	115.2	5.5	1503841	9	US-09-795-668-1	Sequence 1, Appl
C 43	115.2	5.5	1503841	9	US-09-946-807-1	Sequence 1, Appl
C 44	115.2	5.5	1503841	9	US-09-946-807-1	Sequence 1, Appl
C 45	114.4	5.5	611	13	US-10-027-632-204073	Sequence 204073,

ALIGNMENTS

RESULT 1

US-09-974-546-83
; Sequence 83, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert

TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: US/09/974,546
FILING DATE: 10-Oct-2001
CLASSIFICATION: Unknown
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/097,199

/ FILING DATE: 1998-06-12
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 83:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2087 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 99..503
/ SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-974-546-83

Query Match 99.4%; Score 2076; DB 10; Length 2087;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;
Matches 2087; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATATAATTTACAAAATTTATCTTCTA 60
DB 1 GACCTTAAATATATCGAGGTGGCTAATTGATGTATATAATTTACAAAATTTATCTTCTA 60

QY 61 TTGCTACAGAGCTACAATTCAAATTTACAGTAGGCCACCATAGGGGCTTCTTAAGGAACC 120
DB 61 TTGCTACAGAGCTACAATTCAAATTTACAGTAGGCCACCATAGGGGCTTCTTAAGGAACC 120

QY 121 AGAAATATCAGGATATGCAATATTTATTTACAGATTCAGAAATTTAGGSCACA 180
DB 121 AGAAATATCAGGATATGCAATATTTATTTACAGATTCAGAAATTTAGGSCACA 180

QY 181 GATTAAAGTAACCTCCCAAGGCTACAGGCAATCTAGCTCCAGAAATCTGTGCTTTACCAAT 240
DB 181 GATTAAAGTAACCTCCCAAGGCTACAGGCAATCTAGCTCCAGAAATCTGTGCTTTACCAAT 240

QY 241 TCTGCTACAGGTAATTTCCAAAAGGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 300
DB 241 TCTGCTACAGGTAATTTCCAAAAGGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 300

QY 301 TCATTGATTATTCATAGAACAGTACACCATGCAATCTCAGCCCTTGCAGACACACT 360
DB 301 TCATTGATTATTCATAGAACAGTACACCATGCAATCTCAGCCCTTGCAGACACACT 360

QY 361 TGACCATGAAGGTTCCCTCAATGAAATGTTCTCTCAATTTCTTCAAGAGCCATATTATCCA 420
DB 361 TGACCATGAAGGTTCCCTCAATGAAATGTTCTCTCAATTTCTTCAAGAGCCATATTATCCA 420

QY 421 CATTGACTTTGAGTTAACTCAGACCTAGGCTGGAATGCTGCTCTCTACTATTATCCA 480
DB 421 CATTGACTTTGAGTTAACTCAGACCTAGGCTGGAATGCTGCTCTCTACTATTATCCA 480

QY 481 AAATATATACATCACAGATCATATAAATCTCAGCCCTGCTGCAAGAGCCCTTCCAGAAAA 540
DB 481 AAATATATACATCACAGATCATATAAATCTCAGCCCTGCTGCAAGAGCCCTTCCAGAAAA 540

QY 541 ATAAAAATGGTTGAAAGGCAATCTGCTACCAATGACTGTTTAAAGCCAGCCCAAGTAAC 600
DB 541 ATAAAAATGGTTGAAAGGCAATCTGCTACCAATGACTGTTTAAAGCCAGCCCAAGTAAC 600

QY 601 TGAACCATTTCCAACTTCAATTTTATGAAAAGAAATTTGATGATGATGATGATGATGATGAT 660
DB 601 TGAACCATTTCCAACTTCAATTTTATGAAAAGAAATTTGATGATGATGATGATGATGATGAT 660

QY 661 AATTTCTAAATACAAACCATGTTGATCTTTCTCAATCTTGAATCATAGATTTATTTATCT 720
DB 661 AATTTCTAAATACAAACCATGTTGATCTTTCTCAATCTTGAATCATAGATTTATTTATCT 720

QY 721 ATTATCTCAATTTAGTTGTTTATTTATCTAGTGGGCCAATTTAAAACTACCAATGATGTT 780

DB 721 ATTATCTCAATTTAGTTGTTTATTTATCTAGTGGGCCAATTTAAAACTACCAATGATGTT 780

QY 781 TCTGCTCTCCATTTAGTCAATAAATAACGAGCAATTTAGTAAGCAATGTCAGAT 840

DB 781 TCTGCTCTCCATTTAGTCAATAAATAACGAGCAATTTAGTAAGCAATGTCAGAT 840

QY 841 GCTCCGCTAGGCACAGAGGGATAAAAACAATCTTTATAGTATACCACTAAATTTTCGCTT 900

DB 841 GCTCCGCTAGGCACAGAGGGATAAAAACAATCTTTATAGTATACCACTAAATTTTCGCTT 900

QY 901 AGTAACTAGTGAAATGTTCAAGTCTGAGTCAAGAGTTGAGGAGACATTTACAATGT 960

DB 901 AGTAACTAGTGAAATGTTCAAGTCTGAGTCAAGAGTTGAGGAGACATTTACAATGT 960

QY 961 GTAATGGAAAACCAAGGAAAGTGAATCTTTGATAGTGGGACCTAGTGTATTTATATAT 1020

DB 961 GTAATGGAAAACCAAGGAAAGTGAATCTTTGATAGTGGGACCTAGTGTATTTATATAT 1020

QY 1021 TAAATTTGATTTCTGACTCTATCATTTGGCCCTCAAAACACAGATTTGTTTCTTTGTTT 1080

DB 1021 TAAATTTGATTTCTGACTCTATCATTTGGCCCTCAAAACACAGATTTGTTTCTTTGTTT 1080

QY 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCAGTGCCTGCACATAGAAAACAATC 1140

DB 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCAGTGCCTGCACATAGAAAACAATC 1140

QY 1141 AATATTGCTGAATTAATGATTTAAATAATCAGAGAACTTTCCCATTTCTTTGGATCTAT 1200

DB 1141 AATATTGCTGAATTAATGATTTAAATAATCAGAGAACTTTCCCATTTCTTTGGATCTAT 1200

QY 1201 AGAACATCCAGAGTAAGTATGAGGGCTCTGCAATTTATATGCGCTTAAATTTAAGATTAT 1260

DB 1201 AGAACATCCAGAGTAAGTATGAGGGCTCTGCAATTTATATGCGCTTAAATTTAAGATTAT 1260

QY 1261 GTGAGAAAAGTTTAAAGACACTTATAGTAGATTTTGAATATATAGTATAAACAACCTTGGAAA 1320

DB 1261 GTGAGAAAAGTTTAAAGACACTTATAGTAGATTTTGAATATATAGTATAAACAACCTTGGAAA 1320

QY 1321 TGGTGGTCTTTAAAGAGATTTAATAGTAATTAATGAATCTCCATCTCAAAATATATG 1380

DB 1321 TGGTGGTCTTTAAAGAGATTTAATAGTAATTAATGAATCTCCATCTCAAAATATATG 1380

QY 1381 CATAACTATTAAAGGAAATCACTCTCAGGCTTTCAATGTTGTTCTTACTTTT 1440

DB 1381 CATAACTATTAAAGGAAATCACTCTCAGGCTTTCAATGTTGTTCTTACTTTT 1440

QY 1441 CATATATTTTACCATCTGCTGAAGGAGTATATCAAAAGGTAAGAAAGATGGAGGA 1500

DB 1441 CATATATTTTACCATCTGCTGAAGGAGTATATCAAAAGGTAAGAAAGATGGAGGA 1500

QY 1501 AAACTCAGTAAGATTTATTTAGTCTGTTGCAAGTAAAGATTTCTCATCTCAAT 1560

DB 1501 AAACTCAGTAAGATTTATTTAGTCTGTTGCAAGTAAAGATTTCTCATCTCAAT 1560

QY 1561 CCTTATGAGCAGGAGAGGAGGCTGTTTGAAGACCACTTTACTTAGCAGAACCAATAT 1620

DB 1561 CCTTATGAGCAGGAGAGGAGGCTGTTTGAAGACCACTTTACTTAGCAGAACCAATAT 1620

QY 1621 TTTAGACACTTCCCTGCAATTAACATGCAAAACAATATGTTTGCATAAATTTGTTTATCAAC 1680

DB 1621 TTTAGACACTTCCCTGCAATTAACATGCAAAACAATATGTTTGCATAAATTTGTTTATCAAC 1680

QY 1681 CTCCAAACCAACCACTTCAAGGTTAAATTTTCAATCAAACTTCAATGATTTTCTTCTTA 1740

DB 1681 CTCCAAACCAACCACTTCAAGGTTAAATTTTCAATCAAACTTCAATGATTTTCTTCTTA 1740

QY 1741 ACGCTAGAGATTTGCTTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTTAAGAGAGAT 1800

DB 1741 ACGCTAGAGATTTGCTTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTTAAGAGAGAT 1800

QY 1801 TTCTCCCTGTTTAAAGCAGCAAAATTTAGCCATTTCTCTCAAACTTCACTCAATTAAGA 1860

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:58:29 ; Search time 4510.06 Seconds
(without alignments)
17622.418 Million cell updates/sec

Title: US-09-974-546C-83

Perfect score: 2088

Sequence: 1 gacctaataatagaggt.....tactacaacactctacgc 2088

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	119.8	5.7	412	8	AQ045191 RPC111-34
C 2	119	5.7	460	8	AQ829037 HS_5023_B
C 3	116.8	5.6	456	5	BQ353836 PM1-HT034
C 4	116.4	5.6	275	2	BE065285 RC1-BT031
C 5	115.2	5.5	335	5	BQ378810 IL2-UT007
C 6	115.2	5.5	353	2	BQ333793 RC1-HT088
C 7	115.2	5.5	383	1	AA776006 ae82h10.s
C 8	115.2	5.5	396	2	BE155551 PM2-HT035
C 9	115.2	5.5	410	2	BE155550 PM2-HT035
C 10	115.2	5.5	411	2	BE155549 PM2-HT035
C 11	115.2	5.5	416	8	AQ337692 HS_5017_A
C 12	115.2	5.5	466	5	BQ303951 RC1-BT031
C 13	115.2	5.5	474	8	AZ694823 UP_516-21
C 14	115.2	5.5	476	2	BE065326 RC1-BT031
C 15	115.2	5.5	478	2	BE154146 PM1-HT034
C 16	115.2	5.5	480	2	BE065400 RC1-BT031
C 17	115.2	5.5	493	5	BQ303952 RC1-BT031
C 18	115.2	5.5	498	8	B76374 RPC111-15D1
C 19	115.2	5.5	559	8	B54683 CIT-HSP-384
C 20	115.2	5.5	563	2	AW605861 NR0-HT024
C 21	115.2	5.5	573	5	BQ317875 RC0-CT047
C 22	115.2	5.5	638	1	AL698542 DKF2P686D
C 23	115.2	5.5	645	2	BE065286 RC1-BT031
C 24	115.2	5.5	668	2	BE065409 RC1-BT031

25	115.2	5.5	670	1	AI133053
26	115.2	5.5	687	2	BE065399
27	115.2	5.5	703	8	B66360 CIT-HSP-200
28	115.2	5.5	771	2	BE065219 RC1-BT031
29	115.2	5.5	790	1	AU137403
30	115.2	5.5	790	8	BZ600251 WHABAL7TF
31	115.2	5.5	805	8	BZ604730 WHADP83TR
32	114.6	5.5	544	8	AQ600990 HS_5374_A
33	114.6	5.5	724	8	AQ355228
34	114.6	5.5	796	8	BZ611286 WHACL88TR
35	114	5.5	755	9	AG031791 Pan st037
36	113.8	5.5	213	2	AW818026 CM3-ST027
37	113.8	5.5	582	5	BX504526 DKF2P686M
38	113.8	5.5	800	8	BZ600022 WHABK34TF
39	113.6	5.4	270	5	BQ358592 PM1-HT034
40	113.6	5.4	365	2	AW938273 QVO-DT004
41	113.6	5.4	374	2	AW938274 QVO-DT004
42	113.6	5.4	391	2	BE154240 PM1-HT034
43	113.6	5.4	404	8	AQ488905 RPC1-11-2
44	113.6	5.4	415	5	BQ353785 PM1-HT034
45	113.6	5.4	423	1	AV732317

ALIGNMENTS

RESULT 1
AQ045191/c
LOCUS RPC111-34D16.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-34D16,
DEFINITION genomic survey sequence.
ACCESSION AQ045191
VERSION AQ045191.1 GI:3314118
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 412)
AUTHORS Adams,M.D., Rounalev,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
Venter,J.C.
TITLE Use of BAC End Sequences for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Class: BAC ends.

FEATURES

source
1..412
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7512759"
/db_xref="taxon:9606"
/clone="RPC1-11-34D16"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPC1-11"
/note="Vector: pBACel.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

ORIGIN

Query Match 5.7%; Score 119.8; DB 8; Length 412;

Qy	1904	AGTTAAGAAACATCAAAACATAACATCTCGAAGAAACTTCCCAAGTGTAGACTCTGCCTGC	1963
Db	360	ATTAAATGAATCCAGGAGCTGGTCTTTTGAAGAGATCAACATAAATTTGATAGACTGTCTAGC	301
Qy	1964	ACGACACACATATAAAAAAGNAGNGAGAAATCAAATAGACACAAATAAAAAATGATAAAGG	2023
Db	300	AGACTTAATAAGAAANAAGAGAGAAGATCAAATAGACACAAATAANAATGATAAAGG	241
Qy	2024	GATATCACCACCGATCCACAGAAAATACAAACTTACCATCAGAGAATACTACAAACACCTC	2083
Db	240	GATATCACCACCGATCCACAGAAATACAAACTTACCATCAGAGAATACTACAAACACCTC	181
Qy	2084	TACGC	2088
Db	180	TACGC	176

RESULT 3	EST 20-MAY-2002
BQ353836	
LOCUS	456 bp mRNA linear
DEFINITION	PM1-HT0340-150900-013-cl12 HT0340 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BQ353836
VERSION	BQ353836.1 GI:21017892
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 456)
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jorgensen, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE	Simpson, A.O.: Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
PUBMED	10737800
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-HT0340-150900-013-cl1&t&t=2000-09-15&t&t=1) Seq primer: puc 18 forward High quality sequence start: 23 High quality sequence stop: 456.

FEATURES	source
Location/Qualifiers	
1..456	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/dev_stage="Adult"	
/clone_lib="HT0340"	
/note="Organ: head neck; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	

ORIGIN

Query Match	5.6%;	Score 116.8;	DB 5;	Length 456;
Best Local Similarity	98.3%;	Pred. No. 3.2e-15;		

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:07:44 ; Search time 2414 Seconds
(without alignments)
2709.799 Million cell updates/sec

Title: US-09-974-546C-84

Perfect score: 694

Sequence: 1 MFAFLRNQKYMHNIIHL.....TLGLSCLLYLSTKTHPQII 135

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delopt 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cn2_1/USPTO.spool/US09974546/runat_07092005.174502.20945/app_query.fasta.1.654
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPELU=0 -LOOPEXT=0
-UNIT5=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CN 1.1.2647 @runat_07092005.174502.20945 -NCFU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ste.*
12: gb.sv.*
13: gb.un.*
14: gb.vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	2087	6	AR146835 Sequence
2	694	100.0	2103	9	AF189270 Homo sapi
3	694	100.0	2496	9	BC069109 Homo sapi
4	694	100.0	2505	6	AR146836 Sequence

5	694	100.0	2512	9	AF189269 Homo sapi
6	694	100.0	129227	9	HS171N11
7	694	100.0	186698	2	AC036236 Homo sapi
8	99.5	14.3	259920	2	AC096345 Rattus no
9	95	13.7	9833	9	AF291597 Homo sapi
10	94	13.5	9432	9	AF291608 Homo sapi
11	94	13.5	9817	9	AF291631 Homo sapi
12	94	13.5	9817	9	AF291632 Homo sapi
13	94	13.5	9830	9	AF291644 Homo sapi
14	94	13.5	9833	9	AF291587 Homo sapi
15	94	13.5	9833	9	AF291588 Homo sapi
16	94	13.5	9833	9	AF291589 Homo sapi
17	94	13.5	9833	9	AF291590 Homo sapi
18	94	13.5	9833	9	AF291591 Homo sapi
19	94	13.5	9833	9	AF291592 Homo sapi
20	94	13.5	9833	9	AF291593 Homo sapi
21	94	13.5	9833	9	AF291594 Homo sapi
22	94	13.5	9833	9	AF291595 Homo sapi
23	94	13.5	9833	9	AF291598 Homo sapi
24	94	13.5	9833	9	AF291599 Homo sapi
25	94	13.5	9833	9	AF291600 Homo sapi
26	94	13.5	9833	9	AF291601 Homo sapi
27	94	13.5	9833	9	AF291602 Homo sapi
28	94	13.5	9833	9	AF291603 Homo sapi
29	94	13.5	9833	9	AF291604 Homo sapi
30	94	13.5	9833	9	AF291605 Homo sapi
31	94	13.5	9833	9	AF291606 Homo sapi
32	94	13.5	9833	9	AF291607 Homo sapi
33	94	13.5	9833	9	AF291609 Homo sapi
34	94	13.5	9833	9	AF291610 Homo sapi
35	94	13.5	9833	9	AF291612 Homo sapi
36	94	13.5	9833	9	AF291614 Homo sapi
37	94	13.5	9833	9	AF291615 Homo sapi
38	94	13.5	9833	9	AF291618 Homo sapi
39	94	13.5	9833	9	AF291620 Homo sapi
40	94	13.5	9833	9	AF291621 Homo sapi
41	94	13.5	9833	9	AF291623 Homo sapi
42	94	13.5	9833	9	AF291624 Homo sapi
43	94	13.5	9833	9	AF291625 Homo sapi
44	94	13.5	9833	9	AF291626 Homo sapi
45	94	13.5	9833	9	AF291627 Homo sapi

ALIGNMENTS

RESULT 1
AR146835
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

AR146835
Sequence 83 from patent US 6218529.
AR146835
AR146835.1 GI:15110024
Unknown.
Unknown.
Unclassified.
1 (bases 1 to 2087)
An.G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
Biomarkers and targets for diagnosis, prognosis and management of prostate, breast and bladder cancer
Patent: US 6218529-A 83 17-APR-2001;
Location/Qualifiers
1..2087
/organism="unknown"
/mol_type="unassigned DNA"

Alignment Scores:
Pred. No.: 2.18e-67 Length: 2087
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

```
US-09-974-546C-84 (1-135) x AR146835 (1-2087)
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCCTTTTAAGGAACACAGAAATATGAGGATATGCACAATATATTTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLysSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAAATTGAGGCACAGATTAAAGTAACTTCCCAAGGCTACCGAGCATTTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60
Db 219 CCAGAAACTGCTCTTACCATTCGCTACAGGTTAATTCGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 279 AGAAGTCAAAAGGCACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
Db 339 CTCACACCCCTTGACAGACACATTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 399 TCTTCAGAAAGCCATATTATTCATCATTTGACTTTGCGAGTTAACTCAGACCCCTAGGCTCGGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 459 TGCTGTCTTCTCTACTTATCCAAACTATATACATCCACAGATCATTA 503

RESULT 2
AF189270 2103 bp mRNA linear PRI 22-MAR-2001
LOCUS Homo sapiens UC28 protein (UC28) mRNA, complete cds.
DEFINITION AF189270
ACCESSION AF189270.1 GI:1041603
VERSION AF189270.1
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2103)
AUTHORS An, G., Ng, A.Y., Meka, C.S., Luo, G., Bright, S.P., Cazares, L.,
Wright, G.L. Jr. and Veltre, R.W.
TITLE Cloning and characterization of UROC28, a novel gene overexpressed
in prostate, breast, and bladder cancers
JOURNAL Cancer Res. 60 (24), 7014-7020 (2000)
MEDLINE 21028101
PUBMED 11156405
REFERENCE 2 (bases 1 to 2103)
AUTHORS An, G. and Veltre, R.W.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
Parkway, Oklahoma City, OK 73104, USA
FEATURES
source
1..2103
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q23-q24"
/tissue_type="prostate cancer"
1..2103
/gene="UC28"
99..506
/gene="UC28"
/codon_start=1
/product="UC28 protein"
/protein_id="AAG17118.1"
/db_xref="GI:1041604"
/translation="WRAFLRNQKDYEDMNIITHIQIKLRHRLSNFPLDGLIPATPV
LLPFCYKVRKKEKVRKSRQKATFEFDYSIEQSHHAILTPLQTHLTMRKSGMKCSLSLSS

misc_feature
1961..2103
/gene="UC28"
/note="alternative splicing"
2047..2052
/gene="UC28"

polyA_signal
ORIGIN
Alignment Scores: 2,2e-67 Length: 2103
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-974-546C-84 (1-135) x AF189270 (1-2103)
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCCTTTTAAGGAACACAGAAATATGAGGATATGCACAATATATTTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLysSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAAATTGAGGCACAGATTAAAGTAACTTCCCAAGGCTACCGAGCATTTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60
Db 219 CCAGAAACTGCTCTTACCATTCGCTACAGGTTAATTCGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 279 AGAAGTCAAAAGGCACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
Db 339 CTCACACCCCTTGACAGACACATTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 399 TCTTCAGAAAGCCATATTATTCATCATTTGACTTTGCGAGTTAACTCAGACCCCTAGGCTCGGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 459 TGCTGTCTTCTCTACTTATCCAAACTATATACATCCACAGATCATTA 503

REFERENCE 3
BC069109 2496 bp mRNA linear PRI 23-JUN-2004
LOCUS BC069109
DEFINITION Homo sapiens prostate and breast cancer overexpressed 1, mRNA (cdna
clone IMAGE:7216926).
ACCESSION BC069109
VERSION BC069109.1 GI:46575743
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2496)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.E., Brownstein, M.J., Udén, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
```

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 2496)
 Strausberg, R.
 Direct Submission
 Submitted (16-APR-2004) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapsb-rc@mail.nih.gov
 Tissue Procurement: Anup Madan, University of Iowa
 cDNA Library Preparation: Anup Madan, University of Iowa
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Neurogenomics Research Lab,
 200 B EMRB, University of Iowa, Iowa City, IA-52242
anup-madan@uiowa.edu
 Jessica Fahey, Tim Nelson, Jae Goon Yoon and Anup Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: Plate: Row: Column: 0.

FEATURES

source

1. 2496
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:7216926"
 /tissue type="Synthetic constructs"
 /clone_lib="NIH_MGC_242"
 /lab_host="XL10"
 /note="Vector: pUC19"

ORIGIN

Alignment Scores:
 Pred. No.: 2,666-67 Length: 2496
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-974-546C-84 (1-135) x BC069109 (1-2496)

QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIlelleHisIleLeu 20
 Db 96 ATCAGGGCCCTCTTAAGGAACCAAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 155
 QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 Db 156 CAGATCAGAAAATTGAGGCACAGATTAAAGTAACTTCCCAAGGCTACAGGCATTCAGCT 215
 QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
 Db 216 CCAGAACTGTGCTCTTACCATTTCTGCTACAGGATTTTCCGAAAAGAAAGTAAATA 275
 QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 Db 276 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 335
 QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 Db 336 CTCACACCCCTTGACAGACACATTTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 395
 QY 101 SerSerGluAlaIleLeuPheThrLeuGluLeuThrLeuGlnThrLeuGlyLeuGlu 120

Db 396 TCTTCAGAGCCATATATTATTCACATTTGACATTTTGCAGTTAACTCAGACCCCTAGTCTGGAA 455
 QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIlelle 135
 Db 456 TGTGTGTTCTTCTTACTATTCACAAACTATACATATACATCCACAGATCATA 500
 RESULT 4
 ARI46836
 LOCUS ARI46836 2505 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 85 from patent US 6218529.
 ACCESSION ARI46836
 VERSION ARI46836.1 GI:15110025
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 2505)
 AUTHORS An, G., O'Hara, S. Mark., Ralph, D. and Veltri, R.
 TITLE Biomarkers and targets for diagnosis, prognosis and management of
 prostate, breast and bladder cancer
 JOURNAL Patent: US 6218529-A 85 17-APR-2001;
 FEATURES Location/Qualifiers
 1..2505
 /organism="unknown"
 /mol_type="unassigned DNA"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,686-67 Length: 2505
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-09-974-546C-84 (1-135) x ARI46836 (1-2505)
 QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIlelleHisIleLeu 20
 Db 99 ATCAGGGCCCTCTTAAGGAACCAAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
 QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 Db 159 CAGATCAGAAAATTGAGGCACAGATTAAAGTAACTTCCCAAGGCTACAGGCATTCAGCT 218
 QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
 Db 219 CCAGAACTGTGCTCTTACCATTTCTGCTACAGGATTTTCCGAAAAGAAAGTAAATA 278
 QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 Db 279 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
 QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 Db 339 CTCACACCCCTTGACAGACACATTTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 398
 QY 101 SerSerGluAlaIleLeuPheThrLeuLeuGlnLeuThrLeuGlnThrLeuGlyLeuGlu 120
 Db 399 TCTTCAGAGCCATATATTATTCACATTTGACATTTTGCAGTTAACTCAGACCCCTAGTCTGGAA 458
 QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIlelle 135
 Db 459 TGTGTGTTCTTCTTACTATTCACAAACTATACATATACATCCACAGATCATA 503
 RESULT 5
 AFI89269
 LOCUS AFI89269 2512 bp mRNA linear PRI 22-MAR-2001
 DEFINITION Homo sapiens UC28 protein (UC28) mRNA, complete cds.
 ACCESSION AFI89269
 VERSION AFI89269.1 GI:10441601
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2512)
 AUTHORS An.G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
 Wright,G.L. Jr. and Veltri,R.W.
 TITLE Cloning and characterization of UROC28, a novel gene overexpressed
 in prostate, breast, and bladder cancers
 JOURNAL Cancer Res. 60 (24), 7014-7020 (2000)

MEDLINE 21028101
 PUBMED 11156405

REFERENCE 2 (bases 1 to 2512)
 AUTHORS An.G. and Veltri,R.W.
 TITLE Direct Submission
 JOURNAL Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
 Parkway, Oklahoma City, OK 73104, USA

FEATURES
 source
 1..2512
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6q23-q24"
 /tissue_type="prostate cancer"
 1..2512
 /gene="UC28"
 99...506
 /gene="UC28"
 /codon_start=1
 /product="UC28 protein"
 /protein_id="AA017117.1"
 /db_xref="GI:10411602"
 /translation="MRAPRNQKVEDMHNIIHQIKLRHLRSLNPRLPGILAPETV
 LIPFCVKPRKKEVKRSOKATEFIDYSTEQSHAILTLQTLHTWKGSMKCSLSL
 EAILFTLQLTQTGLECLLYLSKTHPQII"
 polyA_signal
 2487..2492
 /gene="UC28"

polyA_signal

ORIGIN
 Alignment Scores:
 Pred. No.: 2,68e-67 Length: 2512
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-974-546C-84 (1-135) x AF189269 (1-2512)

QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIlelleHisIleLeu 20
 |||||
 Db 99 ATGAGGCGCTTCTTAAGGAACCAAGATATGAGGATATGCACATATTATTCACATTTTA 158
 |||||

QY 21 GlnIleArgGlyLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 |||||
 Db 159 CAGATCAGAAAATTGAGGCACAGATTAACTCCCAAGGCTACAGGCATTCCTAGCT 218
 |||||

QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgGlyLysGluLysValLys 60
 |||||
 Db 219 CCAGAAACGTGCTCTTACATCTCTCAAGGATTTTCGAAAAAAGAAAGTAAAGTAA 278
 |||||

QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 |||||
 Db 279 AGAAGTCAAAAGCAACAGATTCAATTGATTATTCATAGACAGTCACACCATGCAATT 338
 |||||

QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
 |||||
 Db 339 CTCACACCCCTTCAGACACACTTGACCATGAAGGTTCTCTCAATGAATGTTCTCTCATTA 398
 |||||

QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 |||||
 Db 399 TCTTCAGAGCCATATTATTCATTTGACTTTTGCAAGTTAACTCAGACCCCTAGGCTCGGAA 458
 |||||

QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
 |||||
 Db 459 TGCTGTCTTCTTACTTATCCAAACTATACATCCACAGATCATA 503
 |||||

RESULT 6
 HSI17IN11
 LOCUS
 DEFINITION Human DNA sequence from clone RP1-171N11 on chromosome 6q23,
 complete sequence.
 ACCESSION AL031433.4 GI:4826442
 VERSION AL031433
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 129227)
 Williams,S.
 Direct Submission
 Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On May 13, 1999 this sequence version replaced gi:4582115.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

COMMENT
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the rare
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr6
 RP1-171N11 is from the library RPCL-1 constructed by the group of
 Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pCVPAC2.

FEATURES
 Location/Qualifiers
 1..129227
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="q23"
 /clone="RP1-171N11"
 /clone_lib="RPCL-1"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,19e-65 Length: 129227
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0

```
DB: 9 Gaps: 0
US-09-974-546C-84 (1-135) x HSI17IN11 (1-129227)

QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20
29892 ATGAGGGCCCTCTTAAGGAACCAAGAAATATGAGGATATGCACATATTATTCACATTITA 29951
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
29952 CAGATCAGAAAATTGAGGCACAGATTAAAGTAATCTCCCAAGGCTACCAAGCATCTTAGCT 30011
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
30012 CCAGAAACTGTGCTCTTACCATCTGCTCAAGAGTATTTCGAAAAAAGAAAAAGTAAAA 30071
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIle 80
30072 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 30131
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
30132 CTCACACCCCTTGACAGACACATTCACCATGAAAGTTCTCTCAATGAAATGTTCTCTCATTA 30191
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
30192 TCTTCAGAGCCATATTATTCACATGACTTTCAGTTCATCAATGAAATGTTCTCTCATTA 30251
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
30252 TGTGTCTTCTCTACTTATTCACAAACTATACATCCACAGATCATTA 30296

RESULT 7
AC036236 186698 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 6 clone RP11-133015, WORKING DRAFT
DEFINITION AC036236
SEQUENCE 7 unordered pieces.
ACCESSION AC036236
VERSION AC036236.2 GI:9958166
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
Waterston,R.H.
2 (bases 1 to 186698)
Direct Submission
Submitted (07-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7523932.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0133015
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Consensus quality: 183373 bases at least Q30
Consensus quality: 184432 bases at least Q40
Consensus quality: 185064 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 186098; sum-of-contigs
```

Quality coverage: 5.84 in Q20 bases; agarose-fp
Quality coverage: 5.57 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 4769: contig of 4769 bp in length
* 4770 4869: gap of unknown length
* 4870 11657: contig of 6788 bp in length
* 11658 11757: gap of unknown length
* 11758 28189: contig of 16432 bp in length
* 28190 28289: gap of unknown length
* 28290 53441: contig of 25052 bp in length
* 53442 53441: gap of unknown length
* 53442 102327: contig of 48886 bp in length
* 102328 102427: gap of unknown length
* 102428 182781: contig of 80354 bp in length
* 182782 182882: gap of unknown length
* 182882 186698: contig of 3817 bp in length.

FEATURES

Location/Qualifiers
source
1..186698
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-133015"

misc_feature

1..4769
/note="assembly_name:Contig10"

misc_feature

4870..11657
/note="assembly_name:Contig11"

misc_feature

11758..28189
/note="assembly_name:Contig12"

misc_feature

28290..53341
/note="assembly_name:Contig13"

misc_feature

53442..102327
/note="assembly_name:Contig14"

misc_feature

102428..182781
/note="assembly_name:Contig15"

misc_feature

clone_end:SP6
vector_side:right
182882..186698

misc_feature

/note="assembly_name:Contig9"

ORIGIN

Alignment Scores:
Pred. No.: 3.3e-65 Length: 186698
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-974-546C-84 (1-135) x AC036236 (1-186698)

QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20

Db 159252 ATGAGGGCCCTCTTAAGGAACCAAGAAATATGAGGATATGCACATATTATTCACATTITA 159311

QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Db 159312 CAGATCAGAAAATTGAGGCACAGATTAAAGTAATCTCCCAAGGCTACCAAGCATCTTAGCT 159371

QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60

Db 159372 CCAGAAAACCTGTGCTCTTACCATCTTCGACAGGTTATTCGAAAAAAGAAAAAGTAAAA 159431

QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80

```

Db 159432 AGAAGTCAAAGGACACAGAGTTTCATTGATTTCATAGAAACAGTCACACCATGCAATT 159491
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLeuGlySerSerMetLeuGlyCysSerLeu 100
Db 159492 CTCACACCTTCAGACACACTTGACCATGAAGGTTCTCAATGAAGATGTCCTCATTA 159551
QY 101 SerSerGluAlaLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 159552 TCTTCAGAGCCATATTATTCATTGACTTTGCAGTTAACTCAGACCTTAGTCTGGAA 159611
QY 121 CysCysLeuLeuTyriusSerIysThrIleHisProGlnIlelle 135
Db 159612 TGTGTCTTCTCTACTATCCAAACATATACATCCACAGATCATA 159656

RESULT 8
AC096345/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-18J15, WORKING DRAFT SEQUENCE, 2
ACCESSION AC096345
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 259920)
AUTHORS Muzny, D., Marie, Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Claveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowls, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nair, L.,
Nwokediemen, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasceriak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puafo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivas, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Stelmle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K.,

```

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 259920)
Worley, K. C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 259920)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBYU
Center clone name: CH230-18J15
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 240867 bases at least Q40
Consensus quality: 244203 bases at least Q30
Consensus quality: 245971 bases at least Q20
Estimated insert size: 252807; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 258395: contig of 258395 bp in length
* 258396 258495: gap of unknown length
* 258496 259920: contig of 1425 bp in length.
FEATURES
Location/Qualifiers
1..259920
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-18J15"
1..1739
/note="wgs end extension
clone end: T7"
misc_feature
8054..8862
/note="clone_boundary
clone_end: T7"

```


ACCESSION AF291608
VERSION AF291608.1 GI:10953210
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 9432)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Worldwide DNA sequence variation in a 10-kilobase noncoding region
on human chromosome 22
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE 20481912
REFERENCE 2 (bases 1 to 9432)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas
at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
77030, USA
FEATURES
source Location/Qualifiers
1..9432
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="Berg 26"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
misc_feature 1..9432
/note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.: 9.95 Length: 9432
Score: 94.00 Matches: 42
Percent Similarity: 34.10% Conservative: 17
Best Local Similarity: 24.28% Mismatches: 38
Query Match: 13.54% Indels: 76
DB: 9 Gaps: 5
US-09-974-546C-84 (1-135) x AF291608 (1-9432)
QY 4 PheLeuArgAsnGlnLysTyrrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
Db 2955 TTTATAATTTCTCAAAAATCCTATGAGCGAGGTATGTAATCTATATTTTGCAGATACAG 3014
QY 24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
Db 3015 AAACGTAGTCATAGGCTAAGTAATTTCCCTAGGCTT----- 3050
QY 44 ValLeuLeuProPheCysTyrrLysValPheArgLysLysGluLysValLysArgSerGln 63
Db 3051 GTAGATCTGGGATTT----- 3065
QY 64 LysAlaThrGluPheIleAspTyrrSerIleGluGlnSerHisAlaIleLeuThrPro 83
Db 3066 ---GAACCCGGGTCTTAGACACTAAG-----GCTAGTCCT 3098
QY 84 LeuGlnThrHisLeuThrMetLysGlySerMetLysCysSerSerLeuSerSerGlu 103
Db 3099 TTAACGTACACCTTTACTGCCTCCAAAATATTGCTGAATGCTCTGATATCTGTCAGGA 3158
QY 104 AlaIle-----LeuPheThrLeu 109
Db 3159 GATATTTTCTCTAATAAATAGGATGGAGGTGTGTGTGGGGAGGCTTTCTAGCTTG 3218
QY 110 ThrLeuGlnLeuThrGlnThrLeuGly----- 118
Db 3219 TCATTCCTTAGGCCCGCAGGGTTGGTGGAGATTAGAGCAAAAATCTTTGGGGAGAGA 3278
QY 119 -----Leu 119

Db 3279 AAAATGGAACCATGGTCTGTCTCTGAATTTCCACAGATCCCTTTGGTCCCAACCTC 3338
QY 120 GluCysCysLeuLeuTyrrLeuSerLysThrIleHisPro 132
Db 3339 CAAGTGAAGGCTCATTTTATTCTCTAGCTCTACACCT 3377
RESULT 11
AF291631 9817 bp DNA linear PRI 24-OCT-2000
LOCUS Homo sapiens isolate South African MeJe 22q11.2 noncoding genomic
DEFINITION sequence.
ACCESSION AF291631
VERSION AF291631.1 GI:10953233
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 9817)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Worldwide DNA sequence variation in a 10-kilobase noncoding region
on human chromosome 22
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE 20481912
REFERENCE 2 (bases 1 to 9817)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas
at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
77030, USA
FEATURES
source Location/Qualifiers
1..9817
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="South African MeJe"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
misc_feature 1..9817
/note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.: 10.4 Length: 9817
Score: 94.00 Matches: 42
Percent Similarity: 34.10% Conservative: 17
Best Local Similarity: 24.28% Mismatches: 38
Query Match: 13.54% Indels: 76
DB: 9 Gaps: 5
US-09-974-546C-84 (1-135) x AF291631 (1-9817)
QY 4 PheLeuArgAsnGlnLysTyrrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
Db 2939 TTTATAATTTCTCAAAAATCCTATGAGCGAGGTATGTAATCTATATTTTGCAGATACAG 2998
QY 24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
Db 2999 AAACGTAGTCATAGGCTAAGTAATTTCCCTAGGCTT----- 3034
QY 44 ValLeuLeuProPheCysTyrrLysValPheArgLysLysGluLysValLysArgSerGln 63
Db 3035 GTAGATCTGGGATTT----- 3049
QY 64 LysAlaThrGluPheIleAspTyrrSerIleGluGlnSerHisAlaIleLeuThrPro 83
Db 3050 ---GAACCCGGGTCTTAGACACTAAG-----GCTAGTCCT 3082
QY 84 LeuGlnThrHisLeuThrMetLysGlySerMetLysCysSerSerLeuSerSerGlu 103

```
Db      3083  TTAACATACCCCTTACTGCCTCAAAAATATTGCTGAATGCTCTGATATCCCTGTCAGGA 3142
QY      104  Alaile-----
Db      3143  GATATTTTCTCTAATAAATAGGATGGAGGTGTGTGTGGGGAGCTTTCTAGCTTG 3202
QY      110  ThrLeuGlnLeuThrGlnThrLeuGly-----
Db      3203  TCATTCCTTTATGGCCAGCAGGTGTGTGAGGATTAGAGCCAAAATCTTTTGGGGAAGAGA 3262
QY      119  -----Leu 119
Db      3263  AAAATGGAACCATGCTGCTGCTGCTGAATCTCCACAGATCCCTTTGGTCCCAACCTC 3322
QY      120  GluCysCysLeuLeuTyrrLeuSerLysThrIleHisPro 132
Db      3323  CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCT 3361

RESULT 12
AF291632      9817 bp  DNA  linear  PRI 24-OCT-2000
LOCUS      Homo sapiens isolate South African Kung 22q11.2 noncoding genomic
DEFINITION
ACCESSION  AF291632
VERSION    AF291632.1  GI:10953234
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9817)
AUTHORS     Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
            Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
            Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE     20481912
PUBMED      11005839
REFERENCE   2  (bases 1 to 9817)
AUTHORS     Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
            Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
            Direct Submission
TITLE       Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES
            Location/Qualifiers
            source
            1..9817
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /isolates="South African Kung"
            /db_xref="taxon:9606"
            /chromosome="22"
            /map="22q11.2"
            misc_feature
            1..9817
            /note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.:      10.4      Length:      9817
Score:          94.00     Matches:    42
Percent Similarity: 34.10%  Conservaive: 17
Best Local Similarity: 24.28%  Mismatches: 38
Query Match:    13.54%   Indels:    76
DB:             9        Gaps:      5

US-09-974-546C-84 (1-135) x AF291632 (1-9817)

QY      4  PheLeuArgAsnGlnLysTyrrGluAspMetHisAsnIleleHisIleLeuGlnIleArg 23
Db      2939  TTTATATTTCTCAAAAATCCCTATGAGGAGGTATGTAATCTATATTTTGCAGATACAG 2998
QY      24  LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
```

```
Db      2999  AAACATGAGTCATAGGCTAAGTAATTTCCCTAGGCTT----- 3034
QY      44  ValLeuLeuProPheCysTyrrLysValPheArgLysLysGluLysValLysArgSerGln 63
Db      3035  GTAGATCTGGGATT----- 3049
QY      64  LysAlaThrGluPheIleAspTyrrSerIleGluGlnSerHisAlaIleLeuThrPro 83
Db      3050  ---GAAACCGGGTTCTTAGACACATAAG-----GCTAGCTCT 3082
QY      84  LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
Db      3083  TTAACATACACCTTTTACTGCCTCCAAAATATTGCTGAATGCTCTGATATCCCTGTCAGGA 3142
QY      104  Alaile-----
Db      3143  GATATTTTCTCTAATAAATAGGATGGAGGTGTGTGTGGGGAGCTTTCTAGCTTG 3202
QY      110  ThrLeuGlnLeuThrGlnThrLeuGly-----
Db      3203  TCATTCCTTTATGGCCAGCAGGTGTGTGAGGATTAGAGCCAAAATCTTTTGGGGAAGAGA 3262
QY      119  -----Leu 119
Db      3263  AAAATGGAACCATGCTGCTGCTGCTGAATCTCCACAGATCCCTTTGGTCCCAACCTC 3322
QY      120  GluCysCysLeuLeuTyrrLeuSerLysThrIleHisPro 132
Db      3323  CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCT 3361

RESULT 13
AF291644      9830 bp  DNA  linear  PRI 24-OCT-2000
LOCUS      Homo sapiens isolate Nigerian 9 22q11.2 noncoding genomic sequence.
DEFINITION
ACCESSION  AF291644
VERSION    AF291644.1  GI:10953246
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 9830)
AUTHORS     Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
            Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
            Worldwide DNA sequence variation in a 10-kilobase noncoding region
            on human chromosome 22
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE     20481912
PUBMED      11005839
REFERENCE   2  (bases 1 to 9830)
AUTHORS     Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
            Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
            Direct Submission
TITLE       Submitted (01-AUG-2000) Human Genetics Center, University of Texas
            at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
            77030, USA
FEATURES
            Location/Qualifiers
            source
            1..9830
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /isolates="Nigerian 9"
            /db_xref="taxon:9606"
            /chromosome="22"
            /map="22q11.2"
            misc_feature
            1..9830
            /note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.:      10.4      Length:      9830
Score:          94.00     Matches:    42
Percent Similarity: 34.10%  Conservaive: 17
```


AUTHORS Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
Familo,P., Trexler,M., Pathy,L., Jorde,L.B., Yu,N. and Li,W.-H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas
at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
77030, USA

FEATURES Location/Qualifiers
source 1..9833
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="Chinese 2"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
misc_feature 1..9833
/note="noncoding region"

ORIGIN

Alignment Scores:
Pred. No.: 10.4 Length: 9833
Score: 94.00 Matches: 42
Percent Similarity: 34.10% Conservative: 17
Best Local Similarity: 24.28% Mismatches: 38
Query Match: 13.54% Indels: 76
DB: 9 Gaps: 5

US-09-974-546C-84 (1-135) x AF291588 (1-9833)

```
QY      4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
Db      2955 TTATATAATTCACAAAATCCTATGAGCGAGGTATGTAATCTATATTTTGACAGATACAG 3014

QY      24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
Db      3015 AAACTGAGTCATAGGCTAAGTAATTCCTAGGCTT----- 3050

QY      44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
Db      3051 GTAGATCTGGGATTT----- 3065

QY      64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIleLeuThrPro 83
Db      3066 ---GAAACCGGGTCTTAGACACTAG-----GCTAGTCCT 3098

QY      84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
Db      3099 TTAACCTACACCTTTACTGCTCCAAAAATATGCTGAATGCTCTGATATCCTGTGACGA 3158

QY      104 AlaIle-----LeuPheThrLeu 109
Db      3159 GATATTTTCTCTAATAAATAGGATGGAGGTGTGTGTGGGGAGCTTTCTAGCTTG 3218

QY      110 ThrLeuGlnLeuThrGlnThrLeuGly----- 118
Db      3219 TCATTCCTTAGGGCCAGCAGGTGTGTGAGGATGAGCCAAAATCTTTGGGGAAGAGA 3278

QY      119 -----Leu 119
Db      3279 AAAATGGAACCATGGTCTGTCTCTGAATTCCTCAACAGATCCCTTTGGTCCCAACCTC 3338

QY      120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
Db      3339 CAAGTAGGCTCATTTTATTCTCTCTAGCTACACCT 3377
```

Search completed: September 9, 2005, 14:44:30
Job time : 2501 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:02:29 ; Search time 309 Seconds
(without alignments)
2586.295 Million cell updates/sec

Title: US-09-974-546C-84

Perfect score: 694

Sequence: 1 MRAFLRNQKYEDMENIHIL.....TLGLECCLLYLSKTHPQII 135

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO spo1/US09974546/runat_07092005.174501.20885/app query.fasta_1.654
-DB=N_Geneseq_16Dec04 -QFWT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09974546 @CGN 1.1 468 @runat_07092005.174501.20885 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	2087	4	Aas04000 Biomarker
2	694	100.0	2088	3	Aaz87583 Prostate
3	694	100.0	2505	4	Aas04001 Biomarker
4	694	100.0	2506	3	Aaz87584 Prostate
5	90	13.0	1113	13	Adt45647 Bacterial

6	90	13.0	1146	13	ADS46591	AdS46591 Bacterial
7	89.5	12.9	4019	4	ABL27798	ABL27798 Drosophil
c 8	83.5	12.0	532	3	AAC38626	AAC38626 Arabidops
c 9	83	12.0	146547	8	ABZ80817	ABZ80817 Human pho
c 10	82	11.8	24601	2	AAx13160	AAx13160 Enterococ
c 11	82	11.8	24601	6	ABS98955	ABS98955 Enterococ
c 12	81.5	11.7	133462	13	ABD32622	ABD32622 Mouse can
c 13	81	11.7	381	8	ABX44573	ABX44573 Bovine ES
c 14	81	11.7	110000	12	ADN46845_11	Continuation (12 o
c 15	81	11.7	110000	12	ADN47591_09	Continuation (12 o
c 16	81	11.7	110000	12	ADN46123_11	Continuation (12 o
c 17	81	11.7	110000	12	ADN47209_09	Continuation (12 o
c 18	81	11.7	110000	12	ADN46464_11	Continuation (12 o
c 19	81	11.7	110000	12	ADN47960_09	Continuation (12 o
c 20	80.5	11.6	1093	2	AAT95761	AAT95761 Arabidops
c 21	80.5	11.6	1093	3	AAC65293	AAC65293 Arabidops
c 22	80	11.5	1693	12	ADQ24476	AdQ24476 Human sof
c 23	79.5	11.5	262	7	ADS31066	AdS31066 Human gen
c 24	79.5	11.5	585	5	ABV54205	ABV54205 Human pro
c 25	79.5	11.5	1464	3	AAC47238	AAC47238 Arabidops
c 26	79.5	11.5	6196	2	AAV13168	AAV13168 Complete
c 27	79.5	11.5	6243	2	AAV13841	AAV13841 Complete
c 28	79.5	11.5	6503	2	AAV13169	AAV13169 Complete
c 29	79.5	11.5	7379	2	AAV13176	AAV13176 Complete
c 30	79.5	11.5	8618	2	AAV18741	AAV18741 Complete
c 31	79.5	11.5	8792	2	AAV18745	AAV18745 Complete
c 32	78.5	11.3	386	8	ABX45278	ABX45278 Bovine ES
c 33	78.5	11.3	398	4	AAI89039	AAI89039 Human pol
c 34	78.5	11.3	4065	12	ADQ65017	AdQ65017 Novel hum
c 35	78.5	11.3	26496	5	ABA19696	ABA19696 Human ner
c 36	78	11.2	3328	4	ABL18311	ABL18311 Drosophil
c 37	77.5	11.2	389	8	ABX39152	ABX39152 Bovine ES
c 38	77.5	11.2	567	10	ABX56974	ABX56974 Arabidops
c 39	77.5	11.2	2411	11	ADM02309	ADM02309 Human cdn
c 40	77.5	11.2	2437	12	ADP22463	ADP22463 Sea-squir
c 41	77.5	11.2	2661	4	ABL26351	ABL26351 Drosophil
c 42	77.5	11.2	3435	4	ABL02615	ABL02615 Drosophil
c 43	77.5	11.2	3958	4	ABL07870	ABL07870 Drosophil
c 44	77.5	11.2	4122	4	ABL07868	ABL07868 Drosophil
c 45	77.5	11.2	6010	10	ADP50278	AdP50278 Human PFM

ALIGNMENTS

RESULT 1	
AA504000	
ID	AA504000 standard; cDNA; 2087 BP.
XX	
AC	AA504000;
XX	
DT	29-AUG-2001 (first entry)
XX	
DE	Biomarker UC band 28 #2, used in diagnosis and prognosis of cancer.
XX	
KW	Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW	benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	99..506
FT	/*tag= a
FT	/product= "Prostate cancer marker protein"
XX	
PN	US6218529-B1.
XX	
PD	17-APR-2001.
XX	
PF	12-JUN-1998; 98US-00097199.
XX	
PR	31-JUL-1995; 95US-0001655P.
PR	11-JAN-1996; 96US-0013611P.
PR	31-JUL-1996; 96US-00692787.

```
XX (UROC-) UROCOR INC.
XX An G, O'hara SM, Ralph D, Veltri R;
XX WPI: 2001-289849/30.
XX P-PSDB; AAU02174.
XX New nucleic acids as biomarkers and targets useful for detecting,
XX diagnosing, prognosing, and in developing treatments for prostate, breast
XX and bladder cancer.
XX Claim 2; Col 117-121; 78pp; English.
XX The sequence represents nucleic acid biomarker UC band 28 #2, used in
XX detection of prostate, breast and bladder cancer. Biomarker nucleic acid
XX sequences can be used as hybridisation probes and primers that
XX specifically hybridise to prostate cancer, benign prostatic hyperplasia
XX (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
XX nucleic acid markers can be used to produce antibodies for the detection
XX of prostate, breast or bladder cancer. The nucleic acids can be used as
XX targets for therapeutic intervention in these diseases, in the
XX identification and isolation of full-length gene sequences, including
XX regulatory elements for gene expression, from genomic human DNA
XX libraries, as hybridisation probes for screening genomic human DNA
XX libraries. The kits comprising the nucleic acid sequences are useful for
XX detecting bladder, breast or prostate cancer cells in a biological sample
XX
XX Sequence 2087 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 2.6e-76 Length: 2087
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-09-974-546C-84 (1-135) x AAS04000 (1-2087)
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
DB 99 ATGAGGGCTCTTTAAGAACACAGAAATATGAGGATATGCACAATATTATTTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAenPheProArgLeuProGlyIleLeuAla 40
DB 159 CAGATCAGAAAATTGAGGCACAGATTAGTAATCTTCCCAAGGCTACCAAGGCAATCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysVallys 60
DB 219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAAGGTATTTCGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisIleAlaIle 80
DB 279 AGAAGTCAAAAGGCAACAGAGTTCTATTGATTATTCATAGAACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
DB 339 CTCACACCTTGCAGACACACTTGACCAATGAAGGTTCTCAATGAATGTTCCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
DB 399 TCTTCAGAAGGCATATTATTACATTGACTTTGCAAGTTAACTCAGACCTAGGTCGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
DB 459 TGCTGTCTCTCTACTTATCCAAAACTATACATCCACAGATCAT 503
RESULT 2
AAZ87583
ID AAZ87583 standard; DNA; 2088 BP.
XX
XX AC AAZ87583;
```

```
XX 19-APR-2000 (first entry)
XX Prostate disease marker UC Band #28.
XX Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
XX benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
XX diagnosis; PCR primer; ss.
XX Homo sapiens.
XX WO9964631-A1.
XX 16-DEC-1999.
XX 11-JUN-1999; 99WO-US013151.
XX 12-JUN-1998; 98US-00097199.
XX (UROC-) UROCOR INC.
XX An G, O'hara SM, Ralph D, Veltri RW;
XX WPI: 2000-116557/10.
XX P-PSDB; AAY59295.
XX Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
XX breast and bladder cancer.
XX Claim 1; Page 182-183; 191pp; English.
XX The invention provides nucleic acid markers of prostate, breast and
XX bladder cancer. The markers are indicators of malignant transformation of
XX prostate, breast and bladder tissues and are diagnostic of the potential
XX for metastatic spread of malignant prostate tumours. The nucleic acid can
XX also be used as targets for therapeutic intervention in prostate cancer,
XX benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
XX markers may be used to design specific probes and primers, for the rapid
XX analysis of prostate, bladder or breast biopsy samples. The probes and
XX primers may also be used for in situ hybridization or in situ PCR
XX detection and diagnosis. They may also be used to identify and isolate
XX full length gene sequences from various DNA libraries. Antibodies against
XX the polypeptide products of the markers can be used to treat prostate
XX cancer, bladder cancer or breast cancer. The encoded proteins may be used
XX to detect antibodies. The proteins and antibodies can be used in
XX immunodetection methods for detecting or quantifying the cancers, and for
XX clinical diagnosis of these cancers. The antibodies may also be used for
XX radioimaging to quantify and localize the encoded proteins
XX
XX Sequence 2088 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 1 Other;
XX
Alignment Scores:
Pred. No.: 2.6e-76 Length: 2088
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-09-974-546C-84 (1-135) x AAZ87583 (1-2088)
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
DB 99 ATGAGGGCTCTTTAAGAACACAGAAATATGAGGATATGCACAATATTATTTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAenPheProArgLeuProGlyIleLeuAla 40
DB 159 CAGATCAGAAAATTGAGGCACAGATTAGTAATCTTCCCAAGGCTACCAAGGCAATCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysVallys 60
DB 219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAAGGTATTTCGAAAAAAGAAAAAGTAAAA 278
```


QY 61 ArgSerGlnLysalaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 Db 279 AGAAGTCAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
 QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
 Db 339 CTCACACCTTGCAGACACACTTGACCATGAAAGGTTCCCTCAATGAATGTTCTCATTA 398
 QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 Db 399 TCTTCAGAGCCCATATTATTACATTGACTTTGCAGTTAACTCAGACCCCTAGGTCTGGAA 458
 QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
 Db 459 TGTGTCTTCTCTACTTATCCAAACTATACATCCACAGATCATA 503
 RESULT 3
 AAS04001
 ID AAS04001 standard; cDNA; 2505 BP.
 XX AAS04001;
 AC AAS04001;
 DT 29-AUG-2001 (first entry)
 XX Biomarker UC band 28 #3, used in diagnosis and prognosis of cancer.
 DE Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
 KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
 KW Homo sapiens.
 OS Homo sapiens.
 PH Key Location/Qualifiers
 FT CDS 99..506
 FT /*tag= a
 FT /product= "Prostate cancer marker protein"
 XX
 PN US6218529-B1.
 XX 17-APR-2001.
 PD 12-JUN-1998; 98US-00097199.
 PF 31-JUL-1995; 95US-0001655P.
 PR 11-JAN-1996; 96US-0013611P.
 PR 31-JUL-1996; 96US-00692787.
 XX
 PA (UROC-) UROCOR INC.
 XX
 XX An G, O'hara SM, Ralph D, Veltri R;
 PI WPI; 2001-289849/30.
 DR P-PSDB; AAU02175.
 XX
 XX New nucleic acids as biomarkers and targets useful for detecting,
 PT diagnosing, prognosing, and in developing treatments for prostate, breast
 PT and bladder cancer.
 XX
 XX Claim 2; Col 121-125; 78pp; English.
 PS
 XX The sequence represents nucleic acid biomarker UC band 28 #3, used in
 CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
 CC sequences can be used as hybridisation probes and primers that
 CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
 CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
 CC nucleic acid markers can be used to produce antibodies for the detection
 CC of prostate, breast or bladder cancer. The nucleic acids can be used as
 CC targets for therapeutic intervention in these diseases, in the
 CC identification and isolation of full-length gene sequences, including
 CC regulatory elements for gene expression, from genomic human DNA
 CC libraries, as hybridisation probes for screening genomic human DNA
 CC libraries. The kits comprising the nucleic acid sequences are useful for
 CC detecting bladder, breast or prostate cancer cells in a biological sample

SQ Sequence 2505 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.39e-76 Length: 2505
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-09-974-546C-84 (1-135) x AAS04001 (1-2505)

QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisLeu 20
 Db 99 ATGAGGGCCCTTCTTAAGGAACAGAAATATGAGGATATGCACAATATTATTACATTTTA 158
 QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 Db 159 CAGATCAGAAATAATTGAGGCACAGATTAGTAATCTTCCCAAGGCTACCAGGCATTCTAGCT 218
 QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGlyLysValLys 60
 Db 219 CCAGAAACTGTGCTCTTACCACTTCTGTACAAAGGTATTTTCGAAAAAGAAAAAGTAA 278
 QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 Db 279 AGAAGTCAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
 QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
 Db 339 CTCACACCTTGCAGACACACTTGACCATGAAAGGTTCTCAATGAATGTTCTCTCATTA 398
 QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 Db 399 TCTTCAGAGCCCATATTATTACATTGACTTTGCAGTTAACTCAGACCCCTAGGTCTGGAA 458
 QY 121 CysCysLeuLeuTyrLysThrIleHisProGlnIleIle 135
 Db 459 TGTGTCTTCTCTACTTATCCAAACTATACATCCACAGATCATA 503

RESULT 4

AZ87584
 ID AZ87584 standard; DNA; 2506 BP.
 XX
 AC AZ87584;
 XX
 DT 19-APR-2000 (first entry)
 XX
 DE Prostate disease marker UC Band #28 splice variant.

Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
 KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
 KW diagnosis; PCR primer; ss.

OS Homo sapiens.

XX WO9964631-A1.

PD 16-DEC-1999.

PF 11-JUN-1999; 99WO-US013151.

PR 12-JUN-1998; 98US-00097199.

PA (UROC-) UROCOR INC.

PI An G, O'hara SM, Ralph D, Veltri RW;

XX WPI; 2000-116557/10.

DR P-PSDB; AAY59296.

PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
 PT breast and bladder cancer.

XX Claim 1; Page 184-186; 191pp; English.

XX The invention provides nucleic acid markers of prostate, breast and

XX bladder cancer. The markers are indicators of malignant transformation of

XX prostate, breast and bladder tissues and are diagnostic of the potential

XX for metastatic spread of malignant prostate tumours. The nucleic acid can

XX also be used as targets for therapeutic intervention in prostate cancer,

XX benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The

XX markers may be used to design specific probes and primers, for the rapid

XX analysis of prostate, bladder or breast biopsy samples. The probes and

XX primers may also be used for in situ hybridization or in situ PCR

XX detection and diagnosis. They may also be used to identify and isolate

XX full length gene sequences from various DNA libraries. Antibodies against

XX the polypeptide products of the markers can be used to treat prostate

XX cancer, bladder cancer or breast cancer. The encoded proteins may be used

XX to detect antibodies. The proteins and antibodies can be used in

XX immunodetection methods for detecting or quantifying the cancers, and for

XX clinical diagnosis of these cancers. The antibodies may also be used for

XX radioimaging to quantify and localize the encoded proteins

SQ Sequence 2506 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 3.39e-76 Length: 2506

Score: 694.00 Matches: 135

Percent Similarity: 100.00% Conservativity: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 3 Gaps: 0

US-09-974-546C-84 (1-135) x AAZ87584 (1-2506)

QY 1 MetArgAlaPheLeuArgAenGlnLysTyrGluAapMetHisAsnIleIleHisLeu 20

DB 99 ATGAGGCGCTTCTTAAGGAACAGAAATATGAGGATATGCACATATATTCACATTTA 158

QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

DB 159 CAGATCAGAAATTTGAGGCACAGATTAAGTAACCTCCCAAGGCTACCGAGCATTCAGCT 218

QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60

DB 219 CCAGAAACTGTGCTCTTACCATCTTGCTCAAGGTATTTCCGAAAAAGAAAAAGTAA 278

QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80

DB 279 AGAAGTCAAAGGCAACAGATTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338

QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100

DB 339 CTCACACCTTGCAGACACACTTGACCATGAAAGGTTCCTCAATGAAATGTTCTCATTA 398

QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

DB 399 TCTTCAGAGCCATATATTATTCATTGACTTTGCGATTTAACTCAGACCCCTAGGTCTG 458

QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

DB 459 TCGTGTCTTCTTACTTATTCGAAACTATACATTCACGATCATTA 503

RESULT 5

ADT45647

ID ADT45647 standard; cDNA; 1113 BP.

XX AC ADT45647;

XX 02-DEC-2004 (first entry)

XX Bacterial polynucleotide #20398.

XX Recombinant DNA construct; transformed plant; improved plant property;

XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

KW pathogen tolerance; pest tolerance; plant disease resistance;

KW cell cycle pathway modification; plant growth regulator;

KW homologous recombination; seed oil yield; protein yield; carbohydrate;

KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

KW bacterial polynucleotide; gene; ss.

OS Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide

XX for expression of a polynucleotide encoding a polypeptide from a

XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 44085; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a

XX promoter functional in a plant cell, where the promoter is positioned to

XX provide for expression of a polynucleotide encoding a polypeptide from a

XX microbial source. The invention also relates to a transformed plant

XX comprising the recombinant DNA construct and a method of producing a

XX transformed plant having an improved property. The plant is a crop plant

XX such as maize or soybean. The method of producing a transformed plant

XX having an improved property comprises transforming a plant with the

XX recombinant DNA construct and growing the transformed plant, where the

XX polynucleotide or polypeptide is useful for improving plant properties.

XX The recombinant DNA construct is useful for producing plants with

XX improved plant properties, e.g. improved cold, heat or drought tolerance,

XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,

XX increased resistance to plant disease, better growth rate by modification

XX of the cell cycle pathway with plant growth regulators, increased rate of

XX homologous recombination, modified seed oil or protein yield and/or

XX content, improved yield by modification of carbohydrate, nitrogen or

XX phosphorus use and/or uptake, by modification of photosynthesis or by

XX providing improved plant growth and development under at least one stress

XX condition, improved lignin production or improved galactomannan

XX production. This sequence represents a bacterial polynucleotide used in

XX the scope of the invention. Note: The sequence data for this patent did

XX not form part of the printed specification but was obtained in electronic

XX format from USPTO at seqdata.uspto.gov/sequence.html.

SQ Sequence 1113 BP; 379 A; 204 C; 242 G; 288 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0.53 Length: 1113

Score: 90.00 Matches: 34

Percent Similarity: 44.88% Conservativity: 23

Best Local Similarity: 26.77% Mismatches: 38

Query Match: 12.97% Indels: 32

DB: 13 Gaps: 7

US-09-974-546C-84 (1-135) x ADT45647 (1-1113)

QY 13 MethHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu----- 29

DB 61 ATGCACACTTA-----GCCATAAAGCTGAGAGAGAGAGGCGCATGAGTGGGATATGTC 114

QY 30 ---SerAsnPro----- 33
 Db 115 ACAACACAGGCCCCAGGAAAGAGAGCTTAAGAGATATGGAATAGAGCTCATA 174
 QY 34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
 Db 175 AAGATCCAGGAATTATAAGTCCT-----TTTTAGATGTA 210
 QY 54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheLele---AspTyrSer 72
 Db 211 AATTTAACTTATGATGTAATCTCAGAGAGCTCAACGAATTTCTTGAAGACTTCGAC 270
 QY 73 IleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
 Db 271 ATAATTCATTTCTCATCATGCA---TTCACACCTCTCTCTTAAAGGCTTAAAGCTGGA 327
 QY 93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
 Db 328 AAGAATATGGAAGGAGGACCTGCTTAACAACACTCACAGCATTTCTTGGCCCATGAATCA 387
 QY 112 GlnLeuThrGlnThrLeuGly 118
 Db 388 AAGCTCTGGGATACTTTGGGG 408
 RESULT 6
 ADS46591
 ID ADS46591 standard; cDNA; 1146 BP.
 XX AC
 AC ADS46591;
 XX DT
 DT 02-DEC-2004 (first entry)
 XX DE
 DE Bacterial polynucleotide #1334.
 XX KW
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.
 XX KW
 OS Bacteria.
 XX PN
 PN US2003233675-A1.
 XX PD
 PD 18-DEC-2003.
 XX PF
 PF 20-FEB-2003; 2003US-00369493.
 XX PR
 PR 21-FEB-2002; 2002US-0360039P.
 XX PA
 PA (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX PI
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 XX FT
 FT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX PS
 PS Claim 1; SEQ ID NO 25021; 122pp; English.
 XX CC
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a

CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX SQ
 SQ Sequence 1146 BP; 388 A; 207 C; 253 G; 298 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.553 Length: 1146
 Score: 90.00 Matches: 34
 Percent Similarity: 44.88% Conservative: 23
 Best Local Similarity: 26.77% Mismatches: 38
 Query Match: 12.97% Indels: 32
 DB: 13 Gaps: 7
 US-09-974-546C-84 (1-135) x ADS46591 (1-1146)
 QY 13 MetHisAenIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu----- 29
 Db 70 ATGCACAACTTA-----GCCATAAAGCTGAGAGAGAGAGGCGCATGAGTGGAAATAGTC 123
 QY 30 ---SerAsnPro----- 33
 Db 124 ACAACACACAGCCGCCAGGAAAGAGAGAGCTTAAGAGATATGGAATAGAGCTCATA 183
 QY 34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
 Db 184 AAGATCCAGGAATTATAAGTCCT-----TTTTAGATGTA 219
 QY 54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheLele---AspTyrSer 72
 Db 220 AATTTAACTTATGATGTAATCTCAGAGAGCTCAACGAATTTCTTGAAGACTTCGAC 279
 QY 73 IleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
 Db 280 ATAATTCATTTCTCATCATGCA---TTCACACCTCTCTCTTAAAGGCTTAAAGCTGGA 336
 QY 93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
 Db 337 AAGAATATGGAAGGAGGACCTGCTTAACAACACTCACAGCATTTCTTGGCCCATGAATCA 396
 QY 112 GlnLeuThrGlnThrLeuGly 118
 Db 397 AAGCTCTGGGATACTTTGGGG 417
 RESULT 7
 ABL27798
 ID ABL27798 standard; DNA; 4019 BP.
 XX AC
 AC ABL27798;
 XX DT
 DT 26-MAR-2002 (first entry)
 XX DE
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 34867.
 XX KW
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX

OS Drosophila melanogaster.

XX WO200171042-A2.

XX PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

XX Claim 1; SEQ ID NO 34867; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AB157737-AB172072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4019 BP; 1158 A; 966 C; 927 G; 368 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.93 Length: 4019
Score: 89.50 Matches: 35
Percent Similarity: 39.84% Conservative: 16
Best Local Similarity: 27.34% Mismatches: 40
Query Match: 12.90% Indels: 37
DB: 4 Gaps: 5

US-09-974-546C-84 (1-135) x ABL27798 (1-4019)

QY 18 HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGly 37
DB 1389 CATAGACTAAAGCAGCAAGGTGCAACACCGTAATATATTTTATAGCGTGTCCAGCG 1448
QY 38 IleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGlu 57
DB 1449 CATAGCAGCCCATCACA-----TCCATCGCATCAAGAACGAAGAAAGCA 1499
QY 58 LysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluInSerHis 77
DB 1500 CGTCAGCAGCAGGAAACTAAGTGTACTTCGAAGATA-----CAT 1538
QY 78 HisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCys 97
DB 1539 TGTACAAATTTACATGCGCGTG-----AGGTGC 1565
QY 98 SerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnThrLeu 117
DB 1566 ACCTCGAAACCTTAATTTGGTTTATAGCCTCATTAAGCTTGCAGCTAGAAGATGTGG 1625
QY 118 GlyLeuGlu-----CysCys----- 122
DB 1626 GGTCAAGAATAGCATCAATAATTTGCTCATACCGTGTGTACATCGCATTAGTGTAGG 1685
QY 123 ---LeuLeuTyrLeuSerLysThr 129
DB 1686 GAAGCGATGTATCTCCCTCTTACA 1709

RESULT 8

AAC38626/c

ID AAC38626 standard; DNA; 532 BP.

XX AAC38626;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 21658.

XX Hybridisation assay; Genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-01343768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137503P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.


```
Db 385 TATAACTTACCAATCTAGTTGGTGTTCACACGGAACA-----TGT 338
QY 50 TyrLysValPheArgLysLysGluLysValLysAArgSerGlnLysAlaThrGluPhe--- 68
Db 337 AGCCATGGGTGCTCTTCCAGAAAGGTAGAGCGCTCTTTCATGCGCCTCCCGAGTTGTA 278
QY 69 -----IleAspTyrSerIleGluGlnSerHisHisAla-----IleLeuThr--- 82
Db 277 GTTTTGTGTCATACTTACGATCTCACTCGCTGCAACGCTGCAGATGCTGACCGGC 218
QY 83 -----ProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSer 98
Db 217 CAACCGGTGAACCCAGCATCATCTGACCGCTCCATTTCGCCAGTACCTCGTGTCTC 158
QY 99 SerLeu-----SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGln 115
Db 157 TCTACTCTCTGACTCTCCACATGTCATCATGTTTCACTATG----- 116
QY 116 ThrLeuGlyLeuGluCysCys-----LeuLeuTyrLeuSer 127
Db 115 TCTCTTGCTACACAGTGTGTTCTGCGCTGATCTCTGCTTATCA 71
RESULT 9
ABZ80817/c
ID ABZ80817 standard; DNA; 146547 BP.
XX
AC ABZ80817;
XX
DT 13-JUN-2003 (first entry)
XX
DE Human phospholipase C gamma 1 polymorphism G329ul.
XX
KW human; single nucleotide polymorphism; SNP; phospholipase C gamma 1;
KW PLCG1; Gene; vascular disease; plasminogen activator inhibitor type 2;
KW PAI-2; diagnosis; atherosclerosis; coronary artery disease; ischemia;
KW myocardial infarction; stroke; thromboembolism; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(64001,c)
FT /*tag= a
XX
PN WO2003007801-A2.
XX
PD 30-JAN-2003.
XX
PF 19-JUL-2002; 2002WO-US023041.
XX
PR 20-JUL-2001; 2001US-0306941P.
PR 28-AUG-2001; 2001US-0315572P.
PR 05-OCT-2001; 2001US-0327488P.
PR 14-DEC-2001; 2001US-00017128.
XX
PA (VITI-) VITIVITY INC.
XX
PI McCarthy J;
XX
XX WPI; 2003-278312/27.
XX
DR
PT Diagnosing and prognosing vascular disease, e.g. coronary artery disease
PT and myocardial infarction, based on the detection of polymorphisms in the
PT phospholipase C gamma 1 and plasminogen activator inhibitor type 2 genes
PT and polypeptides.
XX
PS Claim 44; Page 204-245; 295pp; English.
XX
CC The invention relates to methods, agents and apparatus for diagnosing and
CC prognosing vascular diseases based on the detection of polymorphisms in
CC the phospholipase C gamma 1 (PLCG1) and plasminogen activator inhibitor
CC type 2 (PAI-2) genes and polypeptides. In a preferred claim of the
CC invention the polymorphism is a C to T nucleotide change at position
```

```
CC 64001 of the PLCG1 gene (reference sequence GI 11345540) causing an amino
CC acid change from Threonine to Isoleucine at position 813 of the protein.
CC This sequence represents the region of the PLCG1 gene with the T form of
CC the polymorphism. The methods, agents and apparatus is used for
CC diagnosing and prognosing vascular diseases e.g. atherosclerosis,
CC coronary artery disease (especially), myocardial infarction (especially),
CC ischemia, stroke, peripheral vascular diseases, venous thromboembolism
CC and pulmonary embolism
XX
SQ Sequence 146547 BP; 39552 A; 32625 C; 33339 G; 41031 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.69e+03 Length: 146547
Score: 83.00 Matches: 19
Percent Similarity: 65.71% Conservative: 4
Best Local Similarity: 54.29% Mismatches: 10
Query Match: 11.96% Indels: 2
DB: Gaps: 1

US-09-974-546C-84 (1-135) x ABZ80817 (1-146547)
QY 16 IleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeu 35
Db 29020 ATTATCCCATTTTCCAGATGAAATACTGAGGCACAGAGCTAAAGCAGGATTCAAAC 28961
QY 36 ProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyr 50
Db 28960 CCAGGTAGTCTGGCCCCAGAGTCAGTGCTG-----TTTTCGCTTC 28922

RESULT 10
AAAX13160/c
ID AAAX13160 standard; DNA; 24601 BP.
XX
AC AAAX13160;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:223.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN WO9850555-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US008985.
XX
PR 06-MAY-1997; 97US-0044031P.
PR 16-MAY-1997; 97US-0046655P.
PR 14-NOV-1997; 97US-0066009P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Dillon PJ, Barash SC;
XX
XX WPI; 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides -
PT used to develop products for the detection of Enterococcus and for use in
PT vaccines for prevention or attenuation of Enterococcus infection.
XX
PS Claim 1; Page 1125-1137; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAAX12938 to AAAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence of
CC Enterococcus faecalis in samples. They can also be used for diagnosing
```

CC Enterococcal infection in an animal and monitoring progression of
 CC disease, and for identifying agents which can be used to modulate the
 CC growth or pathogenicity of Enterococcus faecalis, or another related
 CC organism, in vivo or in vitro. In particular the polypeptides encoded by
 CC the Enterococcus faecalis nucleotide sequences can be used in vaccines to
 CC prevent or attenuate an Enterococcal infection
 XX
 SQ Sequence 24601 BP; 8032 A; 4435 C; 5259 G; 6859 T; 0 U; 16 Other;

Alignment Scores:
 Pred. No.: 472 Length: 24601
 Score: 82.00 Matches: 42
 Percent Similarity: 47.10% Conservative: 23
 Best Local Similarity: 30.43% Mismatches: 51
 Query Match: 11.82% Indels: 22
 DB: 2 Gaps: 8

US-09-974-546C-84 (1-135) x AAX13160 (1-24601)
 QY 7 AenGlnLysTyrGluasp-----MetHisAsnIlelle----- 17
 Db 2571 AACCAAAATCAAAATCAACTGAACTTTTGAAGACATTTTTCACGACTTAATAAACAAC 2512
 QY 18 -----HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPhe 32
 Db 2511 GCTTAAATTAAGCACAATTCATCAAAAGAGCCAGTAAATAATAATAAGCCACCCGTT 2452
 QY 33 ProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysVal 52
 Db 2451 CCATCATTAATAATAATAATTCGACCACACCGCCACTCTCTTAACTATATACTATA 2392
 QY 53 PheArgLysGlnLysValLysArgSerGlnLysAlaThrGluPheLeuSerTyrSer 72
 Db 2391 AATCAAGAAAGAAAGAAATGAAGCTTCTCAAAA-----AGAGGCTTCATT---TTTCT 2338
 QY 73 IleGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
 Db 2337 ATT---CATTAAACCGCAGGAATATTGGCGCCTTTAGAAATTTTCTTGATGATTAC 2281
 QY 93 SerSerMetLysCysSerLeuSerSerGluAlaIleLeuPheThrLeuThr----- 110
 Db 2280 ACTTCTTTTGTGTCATCGCT---TCTTTTAAACCGTTCATAAAA---AGAGGCTTCATT---TTTCT 2338
 QY 111 ---LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSer 127
 Db 2223 CCTTTCGGTGTCA---ACTTGAATGCGTAATGTTTACTTCTTCTTATCT 2173

RESULT 11
 ABS98955/c
 ID ABS98955 standard; DNA; 24601 BP.
 XX
 AC ABS98955;
 XX
 DT 18-DEC-2002 (first entry)
 XX
 DE Enterococcus faecalis contig sequence #233.
 XX
 KW Computer readable medium; Enterococcus faecalis; microbe; growth;
 KW pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
 KW therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
 KW biotech technology; antibacterial; modulator of nucleic acid expression;
 KW contig; ds.
 XX
 OS Enterococcus faecalis.
 XX
 PN US2002120116-A1.
 XX
 PD 29-AUG-2002.
 XX
 PF 04-MAY-1998; 98US-00070927.
 XX
 PR 04-MAY-1998; 98US-00070927.
 XX

PA (KUNS/) KUNSCH C A.
 PA (DILL/) DILLON P J.
 XX (BARA/) BARASH S.
 PI Kunsch CA, Dillon PJ, Barash S;
 XX WPI; 2002-750065/81.
 XX
 PT Computer readable medium having recorded on it a Enterococcus faecalis
 PT nucleotide sequence useful for detecting diseases related to Enterococcus
 PT infections in animals.
 XX
 PS Claim 1; Page; 119pp; English.
 XX
 CC The present invention relates to a new computer readable medium with an
 CC Enterococcus faecalis nucleotide sequence. The invention is useful to
 CC diagnose the presence of E. faecalis in a sample or determining the
 CC presence of a specific microbe in a sample. The invention is also useful
 CC for modulating the growth or pathogenicity of E. faecalis, in a vaccine to
 CC confer resistance to Enterococcal infection, for commercial, therapeutic
 CC and industrial purposes, and for fermenting a particular sugar source or
 CC to produce a particular metabolite. The invention is useful for detecting
 CC diseases related to Enterococcus infections in animals, and for detecting
 CC E. faecalis using biotech technology. The present nucleic acid sequence
 CC represents an Enterococcus faecalis contig DNA sequence of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from USPTO
 CC at http.sequdata.uspto.gov
 XX
 SQ Sequence 24601 BP; 8032 A; 4435 C; 5259 G; 6859 T; 0 U; 16 Other;

Alignment Scores:
 Pred. No.: 472 Length: 24601
 Score: 82.00 Matches: 42
 Percent Similarity: 47.10% Conservative: 23
 Best Local Similarity: 30.43% Mismatches: 51
 Query Match: 11.82% Indels: 22
 DB: 2 Gaps: 8

US-09-974-546C-84 (1-135) x ABS98955 (1-24601)
 QY 7 AenGlnLysTyrGluasp-----MetHisAsnIlelle----- 17
 Db 2571 AACCAAAATCAAAATCAACTGAACTTTTGAAGACATTTTTCACGACTTAATAAACAAC 2512
 QY 18 -----HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPhe 32
 Db 2511 GCTTAAATTAAGCACAATTCATCAAAAGAGCCAGTAAATAATAATAAGCCACCCGTT 2452
 QY 33 ProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysVal 52
 Db 2451 CCATCATTAATAATAATAATTCGACCACACCGCCACTCTCTTAACTATATACTATA 2392
 QY 53 PheArgLysGlnLysValLysArgSerGlnLysAlaThrGluPheLeuSerTyrSer 72
 Db 2391 AATCAAGAAAGAAAGAAATGAAGCTTCTCAAAA---AGAGGCTTCATT---TTTCT 2338
 QY 73 IleGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
 Db 2337 ATT---CATTAAACCGCAGGAATATTGGCGCCTTTAGAAATTTTCTTGATGATTAC 2281
 QY 93 SerSerMetLysCysSerLeuSerSerGluAlaIleLeuPheThrLeuThr----- 110
 Db 2280 ACTTCTTTTGTGTCATCGCT---TCTTTTAAACCGTTCATAAAA---AGAGGCTTCATT---TTTCT 2338
 QY 111 ---LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSer 127
 Db 2223 CCTTTCGGTGTCA---ACTTGAATGCGTAATGTTTACTTCTTCTTATCT 2173

RESULT 12
 ABD32622/c
 ID ABD32622 standard; DNA; 133462 BP.
 XX

XX Claim 2; SEQ ID NO 9738; 245pp; English.

XX The invention relates to a purified nucleic acid molecule associated with

XX lactation or muscle and fat deposition (designated LMFD), derived from

XX cattle, and the LMFD nucleic acid can specifically hybridize to a second

XX nucleic acid molecule comprising any of 15112 nucleotide sequences,

XX appearing as ABX34836-ABX49947, or complements of them. Also included are

XX ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic

XX acid linked to a promoter and a 3' non-translated sequence that

XX functions in the cell to cause termination of transcription and addition

XX of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and

XX (2) determining a level or pattern of a molecule in a bovine cell or

XX tissue comprising: (a) incubating a marker nucleic acid (comprising any

XX of the 15112 nucleic acid sequences or its complement or fragment) with a

XX complementary nucleic acid molecule obtained from the bovine cell or

XX tissue, where hybridisation between the marker nucleic acid and the

XX complementary nucleic acid permits the detection of the molecule; and (b)

XX detecting the level or pattern of the complementary nucleic acid; where

XX the detection of the complementary nucleic acid is predictive of the

XX level or pattern of the molecule. The LMFD nucleic acid is used for

XX determining a level or pattern of a molecule in a bovine cell or tissue.

XX It is useful for genome mapping, gene identification and analysis, cattle

XX breeding, preparation of constructs for use in cattle gene expression, or

XX for genetically improving cattle. The present sequence is one of the

XX 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The

XX present sequence was not shown in the specification but was obtained in

XX electronic format from the USPTO web site:

XX seqdata.uspto.gov/sequence.html?DocID=20020137139

XX

SQ Sequence 381 BP; 104 A; 84 C; 82 G; 111 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.51	Length:	381
Score:	81.00	Matches:	24
Percent Similarity:	42.55%	Conservative:	16
Best Local Similarity:	25.53%	Mismatches:	30
Query Match:	11.67%	Indels:	24
DB:	8	Gaps:	4

US-09-974-546C-84 (1-135) x ABX44573 (1-381)

QY 24 LysLeuArgHisArgLeuSerAsnPro-----ArgLeuProGly----- 37

DB 96 AAGATAGACACAGAGTCTCAAACTGCCAATAGTGTCTGCCAGTAGCTGACGACGG 155

QY 38 -----IleLeuAlaProGluThrValLeuLeuProPheCys 49

DB 156 AAAGGTTTGCAACCGTTCGCCCGGTACACCTGCCCGTTTCAGTGCACCTGCTTCT 215

QY 50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69

DB 216 GCAAGTGTGTACGTACGAGGAGTATGTC-----TTCAAT 251

QY 70 AspTyrSerIleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThr 89

DB 252 GAATTTGTGTCGAGACAGCTCAACATGTTAATGTAGATCTTTAAATCGAAACTTACA 311

QY 90 MetLys-----GlySerSerMetLysCysSerSerLeuSer 101

DB 312 TTCAGTTGTCTTCGAGGAGTATGATAATTGTAAACATTAAAT 353

RESULT 14

ADN46845_11/c

Continuation (12 of 21) of ADN46845 from base 1100001 (Thermococcus kodakaraensis KOD1

WP Sequence split into 21 fragments LOCUS ADN46845 Accession Adn46845

WP	Fragment Name	Begin	End
WP	ADN46845_00	1	110000
WP	ADN46845_01	100001	210000
WP	ADN46845_02	200001	310000
WP	ADN46845_03	300001	410000
WP	ADN46845_04	400001	510000
WP	ADN46845_05	500001	610000

WP	ADN46845_06	600001	710000
WP	ADN46845_07	700001	810000
WP	ADN46845_08	800001	910000
WP	ADN46845_09	900001	1010000
WP	ADN46845_10	1000001	1110000
WP	ADN46845_11	1200001	1210000
WP	ADN46845_12	1200001	1310000
WP	ADN46845_13	1300001	1410000
WP	ADN46845_14	1400001	1510000
WP	ADN46845_15	1500001	1610000
WP	ADN46845_16	1600001	1710000
WP	ADN46845_17	1700001	1810000
WP	ADN46845_18	1800001	1910000
WP	ADN46845_19	1900001	2010000
WP	ADN46845_20	2000001	2089378

Alignment Scores:

Pred. No.:	5.52e+03	Length:	110000
Score:	81.00	Matches:	31
Percent Similarity:	44.36%	Conservative:	28
Best Local Similarity:	23.31%	Mismatches:	42
Query Match:	11.67%	Indels:	32
DB:	12	Gaps:	7

US-09-974-546C-84 (1-135) x ADN46845_11 (1-110000)

QY 13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSer----- 30

DB 68597 ATGCACAGCTT-----GCAATCTATCTGAAAAAGCTCGGCACGACGCTTTCAATAGTG 68544

QY 31 -----AsnPro 33

DB 68543 ACTAATGATCTAAAAACGGGAAAGAGCTTTGAAGAGCTGGGGTAGGATTGGTT 68484

QY 34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53

DB 68483 AAGTCCCGGGTGCATCAGTCGG---GTGCTCGGATAAACATAACCTAC----- 68436

QY 54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe---IleAspTyrSer 72

DB 68435 -----GGGCTGAAGTCGAACAGAGAACTAGGCGAGTTCTCGTGATTGTAT 68388

QY 73 IleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92

DB 68387 GTGTGCCAGCCACACAGCC---TTCACCTCCACTCTCGCTGAAGCGGTTAAGCGCGG 68331

QY 93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111

DB 68330 AGAACTCTCGAAAGGCAACACTCTCACACCCACAGCATATCTTCTCCACGAGTCT 68271

QY 112 GlnLeuThrGlnThrLeuGlyLeuGluCysLysLeu 124

DB 68270 TCCCTATGGAAGCCCTTGGGTGACTTTTCCACTCTC 68232

RESULT 15

ADN47591_09

Continuation (10 of 21) of ADN47591 from base 900001 (Thermococcus kodakaraensis KOD1

WP Sequence split into 21 fragments LOCUS ADN47591 Accession Adn47591

WP	Fragment Name	Begin	End
WP	ADN47591_00	1	110000
WP	ADN47591_01	100001	210000
WP	ADN47591_02	200001	310000
WP	ADN47591_03	300001	410000
WP	ADN47591_04	400001	510000
WP	ADN47591_05	500001	610000
WP	ADN47591_06	600001	710000
WP	ADN47591_07	700001	810000
WP	ADN47591_08	800001	910000
WP	ADN47591_09	900001	1010000
WP	ADN47591_10	1000001	1110000
WP	ADN47591_11	1100001	1210000
WP	ADN47591_12	1200001	1310000
WP	ADN47591_13	1300001	1410000

```
WP ADN47591_14 1400001 1510000
WP ADN47591_15 1500001 1610000
WP ADN47591_16 1600001 1710000
WP ADN47591_17 1700001 1810000
WP ADN47591_18 1800001 1910000
WP ADN47591_19 1900001 2010000
WP ADN47591_20 2000001 2089378

Alignment Scores:
Pred. No.: 5.52e+03 Length: 110000
Score: 81.00 Matches: 31
Percent Similarity: 44.36% Conservative: 28
Best Local Similarity: 23.31% Mismatches: 42
Query Match: 11.67% Indels: 32
DB: 12 Gaps: 7

US-09-974-546c-84 (1-135) x ADN47591_09 (1-110000)

QY 13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSer----- 30
Db 20781 ATGCACGAGTT-----GCAATCTATCTGAAAAAGCTCGGCCACGACGTTTCAATAGTG 20834
QY 31 -----AsnPhePro 33
Db 20835 ACTAATGATCTAAAAACGGGAAAGGAAAGAGCTTGAAGAGCTGGGGGTAGGATTGGTT 20894
QY 34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
Db 20895 AAAGTCCCGGTGTCATAGTCGCG---GTGCTCGGGATAAACATAACCTAC----- 20942
QY 54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe---IleAspTyrSer 72
Db 20943 -----GGGCTGAAGTCGAACAGAGAACTAGGCGAGTTTCTCGTGGATTGTGAT 20990
QY 73 IleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
Db 20991 GTGTCCACGCCACACACGCC---TTCACCTCCACTCTCGCTGAAGCGGTTAAGGCCGGG 21047
QY 93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
Db 21048 AGMACTCTCGAAAGGCAACACTCTCCACACCCACGACATATCCTTCTCCACGAGTCT 21107
QY 112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeu 124
Db 21108 TCCCTATGGAAGGCCCTTGGGTTGACTTTTCCACTCCTC 21146
```

Search completed: September 9, 2005, 13:23:49
Job time : 392 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:12:05 ; Search time 104.5 Seconds
(without alignments)
2113.850 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLNQYEDMNHIL.....TLGLECLLYLSTHPIQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09974546/runat_07092005_174503_21006/app_query.fasta_1.654
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CGN 1 1 77 @runat_07092005_174503_21006 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DIV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	2087	3	US-09-097-199-83
2	694	100.0	2505	3	US-09-097-199-85
3	694	100.0	2506	4	US-09-949-016-3311
4	694	100.0	6507	4	US-09-949-016-15053
5	689	99.3	601	4	US-09-949-016-119365
6	689	99.3	601	4	US-09-949-016-119366
7	480	69.2	601	4	US-09-949-016-119364
8	83.5	12.0	2472	4	US-09-248-796A-148
9	80.5	11.6	1093	3	US-09-186-276B-53
10	80.5	11.6	1093	3	US-08-842-445-53
11	80.5	11.6	1093	3	US-09-186-188B-53
12	80.5	11.6	1093	4	US-09-265-585C-53

c 13	80	11.5	148783	4	US-09-949-016-15729	Sequence 15729, A
c 14	79.5	11.5	262	4	US-09-573-080A-99	Sequence 99, Appl
c 15	79.5	11.5	6196	3	US-08-675-566-5	Sequence 5, Appl
16	79.5	11.5	6243	3	US-08-675-566-14	Sequence 14, Appl
17	79.5	11.5	6503	3	US-08-675-566-6	Sequence 6, Appl
18	79.5	11.5	7379	3	US-08-675-566-13	Sequence 13, Appl
19	79.5	11.5	8618	3	US-08-675-566-21	Sequence 21, Appl
20	79.5	11.5	8792	3	US-08-675-566-25	Sequence 25, Appl
21	78	11.2	761	4	US-09-270-767-1826	Sequence 1826, Ap
22	78	11.2	761	4	US-09-270-767-17108	Sequence 17108, A
23	78	11.2	96878	4	US-09-949-016-12551	Sequence 12551, A
24	78	11.2	462589	4	US-09-949-016-12900	Sequence 12900, A
c 25	78	11.2	476044	4	US-09-949-016-12412	Sequence 12412, A
26	77	11.1	1461	1	US-08-587-670A-1	Sequence 1, Appl
27	77	11.1	1461	3	US-09-061-674-1	Sequence 1, Appl
28	77	11.1	22372	4	US-09-949-016-17459	Sequence 17459, A
29	76	11.0	367	3	US-09-328-111-446	Sequence 446, App
c 30	76	11.0	601	4	US-09-949-016-202994	Sequence 202994, A
31	76	11.0	42672	4	US-09-949-016-17253	Sequence 17253, A
32	76	11.0	42672	4	US-09-949-016-17254	Sequence 17254, A
33	76	11.0	66213	4	US-09-949-016-11803	Sequence 11803, A
34	76	11.0	66213	4	US-09-949-016-16739	Sequence 16739, A
c 35	75.5	10.9	19008	4	US-09-949-016-12923	Sequence 12923, A
c 36	75.5	10.9	160759	4	US-09-949-016-16514	Sequence 16514, A
37	75	10.8	64291	4	US-09-949-016-16278	Sequence 16278, A
38	75	10.8	117410	4	US-09-949-016-12262	Sequence 12262, A
39	74.5	10.7	3915	4	US-09-023-655-1104	Sequence 1104, Ap
c 40	74.5	10.7	5523	4	US-09-949-016-1259	Sequence 1259, Ap
c 41	74.5	10.7	5523	4	US-09-949-016-1014	Sequence 1014, Ap
42	74.5	10.7	27120	4	US-09-949-016-16210	Sequence 16210, A
43	74.5	10.7	35707	4	US-09-949-016-17120	Sequence 17120, A
c 44	74	10.7	601	4	US-09-949-016-157153	Sequence 157153, A
c 45	74	10.7	3143	4	US-09-949-016-856	Sequence 856, App

ALIGNMENTS

RESULT 1
US-09-097-199-83
; Sequence 83, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Velttri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:

```
/ ; TELEPHONE: (512) 418-3000
/ ; TELEFAX: (512) 474-7577
/ ; INFORMATION FOR SEQ ID NO: 83:
/ ; SEQUENCE CHARACTERISTICS:
/ ; LENGTH: 2087 base pairs
/ ; TYPE: nucleic acid
/ ; STRANDEDNESS: single
/ ; TOPOLOGY: linear
/ ; FEATURE:
/ ; NAME/KEY: CDS
/ ; LOCATION: 99..503
/ ;
US-09-097-199-83
Alignment Scores:
Pred. No.: 1-65e-85 Length: 2087
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-097-199-83 (1-2087)
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisLeu 20
DB 99 ATGAGGGCCTTCTTAAGGAACACAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
DB 159 CAGATCAGAAAAATTGAGGCACAGATTAAAGTAATCTCCCAAGGCTACCGGCATTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
DB 219 CCAGAAACTGTGCTCTTACCATCTGCTACAGGATTTCGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
DB 279 AGAAGTCAAAAGGCACAGAGTTCAATTGATTATCCATAGAACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
DB 339 CTCACACCTTGACAGACACATTGACATGAAGGTTCTCAATGAATGTTCCCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
DB 399 TCTTCAGAGCCATATTATTCATTGACTTTGCAAGTAACTCAGACCCCTAGGCTCGGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
DB 459 TGTGTCTTCTCTACTTATCCAAAACATATACATCCACAGATCATATA 503

RESULT 2
US-09-097-199-85
/ ; Sequence 85, Application US/09097199
/ ; Patent No. 6218529
/ ; GENERAL INFORMATION:
/ ; APPLICANT: An, Gang
/ ; APPLICANT: O'Hara, S. Mark
/ ; APPLICANT: Ralph, David
/ ; APPLICANT: Velttri, Robert
/ ; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
/ ; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
/ ; NUMBER OF SEQUENCES: 87
/ ; CORRESPONDENCE ADDRESS:
/ ; ADDRESSEE: Arnold, White & Durkee
/ ; STREET: P.O. Box 4433
/ ; CITY: Houston
/ ; STATE: Texas
/ ; COUNTRY: USA
/ ; ZIP: 77210
/ ; COMPUTER READABLE FORM:
/ ; MEDIUM TYPE: Floppy disk
/ ; COMPUTER: IBM PC compatible
```

```
/ ; OPERATING SYSTEM: PC-DOS/MS-DOS
/ ; SOFTWARE: PatentIn Release #1.0, Version #1.30
/ ; CURRENT APPLICATION DATA:
/ ; APPLICATION NUMBER: US/09/097,199
/ ; FILING DATE:
/ ; CLASSIFICATION:
/ ; PRIOR APPLICATION DATA:
/ ; APPLICATION NUMBER: US 08/692,787
/ ; FILING DATE: 31-JUL-1996
/ ; ATTORNEY/AGENT INFORMATION:
/ ; NAME: Nakashima, Richard A.
/ ; REGISTRATION NUMBER: P-42,023
/ ; REFERENCE/DOCKET NUMBER: UROC:018
/ ; TELECOMMUNICATION INFORMATION:
/ ; TELEPHONE: (512) 418-3000
/ ; TELEFAX: (512) 474-7577
/ ; INFORMATION FOR SEQ ID NO: 85:
/ ; SEQUENCE CHARACTERISTICS:
/ ; LENGTH: 2505 base pairs
/ ; TYPE: nucleic acid
/ ; STRANDEDNESS: single
/ ; TOPOLOGY: linear
/ ; FEATURE:
/ ; NAME/KEY: CDS
/ ; LOCATION: 99..503
/ ;
US-09-097-199-85
Alignment Scores:
Pred. No.: 2-21e-85 Length: 2505
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-097-199-85 (1-2505)
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisLeu 20
DB 99 ATGAGGGCCTTCTTAAGGAACACAGAAATATGAGGATATGCACAATATTATTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
DB 159 CAGATCAGAAAAATTGAGGCACAGATTAAAGTAATCTCCCAAGGCTACCGGCATTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
DB 219 CCAGAAACTGTGCTCTTACCATCTGCTACAGGATTTCGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
DB 279 AGAAGTCAAAAGGCACAGAGTTCAATTGATTATCCATAGAACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
DB 339 CTCACACCTTGACAGACACATTGACATGAAGGTTCTCAATGAATGTTCCCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
DB 399 TCTTCAGAGCCATATTATTCATTGACTTTGCAAGTAACTCAGACCCCTAGGCTCGGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
DB 459 TGTGTCTTCTCTACTTATCCAAAACATATACATCCACAGATCATATA 503

RESULT 3
US-09-949-016-3311
/ ; Sequence 3311, Application US/09949016
/ ; Patent No. 6812339
/ ; GENERAL INFORMATION:
/ ; APPLICANT: VENTER, J. Craig et al.
/ ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```

; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; SOFTWARE: FastSeq for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 207012
; SEQ ID NO 3311
; LENGTH: 2506
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-3311

Alignment Scores:
Pred. No.: 2,21e-85 Length: 2506
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-949-016-3311 (1-2506)

Qy 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 97 ATGAGGGCCCTCTTAAGGAAACAGCAATATAGGATATGCACATATATTATTCACATTTTA 156
Qy 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 157 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAAGCATTCAGCT 216
Qy 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysVallys 60
Db 217 CCAGAAACTGTGCTCTTACCATTCCTGCTACAAGGTATTTCCGAAAAAGAAAGTAAAA 276
Qy 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 277 AGAAGTCAAAAGGCAACAGATTCATGTATTTCCATAGAACAGTCACACCATGCAATT 336
Qy 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
Db 337 CTCACACCCCTGCAGACACATTCACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 396
Qy 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 397 TCTTCAGAAAGCCATATATTATTCATGCTTTGCGAGTTAACTCAGACCCCTAGGCTCGAA 456
Qy 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 457 TGCTGTCTTCTACTTATTCACAAACTATACATCCACAGATCATATA 501

RESULT 4
US-09-949-016-15053
; Sequence 15053, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15053
; LENGTH: 6507
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15053

Alignment Scores:
Pred. No.: 9.96e-85 Length: 6507
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-949-016-15053 (1-6507)

Qy 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 2097 ATGAGGGCCCTCTTAAGGAAACAGCAATATAGGATATGCACATATATTATTCACATTTTA 2156
Qy 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 2157 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAAGCATTCAGCT 2216
Qy 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysVallys 60
Db 2217 CCAGAAACTGTGCTCTTACCATTCCTGCTACAAGGTATTTCCGAAAAAGAAAGTAAAA 2276
Qy 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 2277 AGAAGTCAAAAGGCAACAGATTCATGTATTTCCATAGAACAGTCACACCATGCAATT 2336
Qy 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
Db 2337 CTCACACCCCTGCAGACACATTCACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 2396
Qy 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 2397 TCTTCAGAAAGCCATATATTATTCATGCTTTGCGAGTTAACTCAGACCCCTAGGCTCGAA 2456
Qy 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 2457 TGCTGTCTTCTACTTATTCACAAACTATACATCCACAGATCATATA 2501

RESULT 5
US-09-949-016-119365/c
; Sequence 119365, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119365
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119365

Alignment Scores:
Pred. No.: 1.14e-85 Length: 601
Score: 689.00 Matches: 134
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

```

Percent Similarity: 99.26% Conservative: 0
Best Local Similarity: 99.26% Mismatches: 1
Query Match: 99.28% Indels: 0
DB: 4 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-949-016-119365 (1-601)

```
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20
DB 506 ATGAGGCGCTTCTTAAGGAAACAGAAATATGAGGATATGCCAAATATTATTCACATTTTA 447
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
DB 446 CAGATCAGAAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAAGCATTTCTAGCT 387
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysVallys 60
DB 386 CAGAAAACTGTCTCTTACCATTCGTCTACAAAGGTATTTCCGAAAAAGAAAAAGTAAAA 327
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
DB 326 AGAAGTCAAAAGGCAACAGAGTTCAATGATTATTCATAGACAGTCACACCATGCAATT 267
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerMetLysCysSerSerLeu 100
DB 266 CTCACACCCCTGCAGACACACTTGACCATGAAAGGTTCTCAATGAAATGTTCTCTCATTA 207
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlu 120
DB 206 TCTTCAGAAAGCCATATTATTACATTGACTTTGCAGTTAACTCAGACCCCTAGGTCGGAA 147
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
DB 146 TGCTGTCTTCTACTTATCCAAACTATATACATCCACAGATCATA 102
```

RESULT 6

US-09-949-016-119366/c
; Sequence 119366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119366
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119366

Alignment Scores:
Pred. No.: 1,14e-85 Length: 601
Score: 689.00 Matches: 134
Percent Similarity: 99.26% Conservative: 0
Best Local Similarity: 99.26% Mismatches: 1
Query Match: 99.28% Indels: 0
DB: 4 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-949-016-119366 (1-601)

```
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20
DB 518 ATGAGGCGCTTCTTAAGGAAACAGAAATATGAGGATATGCCAAATATTATTCACATTTTA 459
```

```
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
DB 458 CAGATCAGAAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAAGCATTTCTAGCT 399
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysVallys 60
DB 398 CCAGAAAACTGTCTCTTACCATTCGTCTACAAAGGTATTTCCGAAAAAGAAAAAGTAAAA 339
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
DB 338 AGAAGTCAAAAGGCAACAGAGTTCAATGATTATTCAYAGAACAGTCACACCATGCAATT 279
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerMetLysCysSerSerLeu 100
DB 278 CTCACACCCCTGCAGACACACTTGACCATGAAAGGTTCTCAATGAAATGTTCTCTCATTA 219
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlu 120
DB 218 TCTTCAGAAAGCCATATTATTACATTGACTTTGCAGTTAACTCAGACCCCTAGGTCGGAA 159
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
DB 158 TGCTGTCTTCTACTTATCCAAACTATATACATCCACAGATCATA 114
```

RESULT 7

US-09-949-016-119364/c
; Sequence 119364, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119364
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119364

Alignment Scores:
Pred. No.: 9,72e-57 Length: 601
Score: 480.00 Matches: 92
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 69.16% Indels: 0
DB: 4 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-949-016-119364 (1-601)

```
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20
DB 277 ATGAGGCGCTTCTTAAGGAAACAGAAATATGAGGATATGCCAAATATTATTCACATTTTA 218
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
DB 217 CAGATCAGAAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCAAGCATTTCTAGCT 158
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysVallys 60
DB 157 CCAGAAAACTGTCTTACCATTCGTCTACAAAGGTATTTCCGAAAAAGAAAAAGTAAAA 98
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
```

Db 97 AGAGTCAAAGGCAACAGAGTTCATGTATTCATAGACAGTCACACCATGCAATT 38
Qy 81 LeuThrProLeuGlnThrHisLeuThrMetLysGly 92
Db 37 CTCACACCCCTTGACAGACACACTTGACCATGAAAGGT 2
RESULT 8
US-09-248-796A-148
; Sequence 148, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 148
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-148
Alignment Scores:
Pred. No.: 0.697 Length: 2472
Score: 83.50 Matches: 35
Percent Similarity: 44.17% Conservative: 18
Best Local Similarity: 29.17% Mismatches: 52
Query Match: 12.03% Indels: 15
DB: 4 Gaps: 5
US-09-974-546C-84 (1-135) x US-09-248-796A-148 (1-2472)
Qy 24 LysLeuArgHisArgLeuSerAnPheProArgLeuProGlyLeuAlaProGluThr 43
Db 1634 AAAATCCAAATCTACTACTGATCTCCATGCTATCTTGCTGCTGCGGTGACG 1693
Qy 44 ValLeuLeuProPheCys-----TyrLysValPheArgLysLysGluLysVal 59
Db 1694 ATGATGATCAGCATGATGATGAAGAAGAAGATCAAGCTGATGAAGAAGAAGAGGAAG--- 1750
Qy 60 LysArgSerGlnLysAlaThrGluPheLeuAspTyrSerIleGluGlnSerHisAla 79
Db 1751 AAGATGAAGAAGAAGACGATCCAGTTGGAGATTCACGATAGAGATGATTTTGTACA 1810
Qy 80 Ile-----LeuThrProLeuGlnThrHisLeuThrMetLysGlySerMetLysCys 97
Db 1811 ATATTGCTCTTGTCGACACTACCAACAATGACATATGTTATGAANAATGATCTCGATT 1870
Qy 98 SerSerLeuSer-----SerGluAlaLeuLeuPheThrLeuThrLeu 111
Db 1871 TCCAAGTTATTCCAACCTTTTGGTCAATTGATCACCCTCAATTCGTAATCTCAAAACT 1930
Qy 112 GluLeuThrGln-----ThrLeuGlyLeuGluCysLeuLeuTyrLeuSerLysThr 129
Db 1931 CCTTTGCTAAATGTTACGTAATTCACCAATGTTGTTATTACATGTTGTAACATTACT 1990
RESULT 9
US-09-186-276B-53/c
; Sequence 53, Application US/09186276B
; Patent No. 6388173
; GENERAL INFORMATION:
; APPLICANT: Benfey, Philip
; APPLICANT: DiLaurenzio, Laura
; APPLICANT: Wysocka-Diller, Joanna
; APPLICANT: Malamy, Jocelyn E.
; APPLICANT: Pysh, Leonard

; APPLICANT: Helariutta, Yrjo
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
; FILE REFERENCE: 5914-075-999
; CURRENT APPLICATION NUMBER: US/09/186,276B
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 1093
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-186-276B-53
Alignment Scores:
Pred. No.: 0.5 Length: 1093
Score: 80.50 Matches: 37
Percent Similarity: 36.81% Conservative: 16
Best Local Similarity: 25.69% Mismatches: 47
Query Match: 11.60% Indels: 44
DB: 3 Gaps: 6
US-09-974-546C-84 (1-135) x US-09-186-276B-53 (1-1093)
Qy 10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
Db 921 TACAGCGACAGATGTTTTCCTGCTCTCTTTCCTCACTCAAGTAACCATCATCAC 862
Qy 30 SerAsnPheProArgLeuProGlyLeuAlaProGluThrValLeuLeuProPheCys 49
Db 861 TATAACTTACCAATATATGTTGGGTTTGGCTTCCACACGGAACA-----TGT 814
Qy 50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
Db 813 ACCCATGGTCTCTCTTCCAGAGAGGATGAGCGCTCTTCATGCGCTCCCGATTGTA 754
Qy 70 AspTyrSerIle-----GluGlnSerHisAlaLeuLeuThrProLeuGlnThrHis-LeuThrMe 73
Db 753 GTTTTGTGTCATAAGCTTTTCAGCATCTCACTCGCTCAAAACGCTGCAGATGCTGACCGG 694
Qy 74 -----GluGlnSerHisAlaLeuLeuThrProLeuGlnThrHis-LeuThrMe 90
Db 693 CCAACCCCGTGAACCCAGCCATCAT-----CATTTGACCTT----- 658
Qy 90 tLysGlySerSerMetLysCysSerSerLeu-----SerSerGluAlaIleLeuPh 107
Db 657 CCATTTCCCACTACCTCGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 598
Qy 107 eThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCys-----LeuLe 124
Db 597 CACTATG-----TCTTTGCTACACAGTGTGTTCTGCGCTGATTCT 556
Qy 124 uTyrLeuSer 127
Db 555 CTGCTTATCA 546
RESULT 10
US-08-842-445-53/c
; Sequence 53, Application US/08842445A
; Patent No. 6441270
; GENERAL INFORMATION:
; APPLICANT: Benfey et al
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
; FILE REFERENCE: 5914-056-999
; CURRENT APPLICATION NUMBER: US/08/842,445A
; PRIOR FILING DATE: 1997-04-24
; EARLIER APPLICATION NUMBER: 08/638,617
; EARLIER FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79

/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 53
/ LENGTH: 1093
/ TYPE: DNA
/ ORGANISM: Plant
US-08-842-445-53

Alignment Scores:
Pred. No.: 0.5 Length: 1093
Score: 80.50 Matches: 37
Percent Similarity: 36.81% Conservative: 16
Best Local Similarity: 25.69% Mismatches: 44
Query Match: 11.60% Indels: 47
DB: 3 Gaps: 6

US-09-974-546C-84 (1-135) x US-08-842-445-53 (1-1093)

```
QY 10 TyrGluAspMetHisAsnIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
Db 921 TACAGCGACAGATGTGTTTGTGCTCTCTTTATCCACTCAAGTAACCATCATCAC 862
QY 30 SerAenPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
Db 861 TATACTTACCCAAATATAGTTTGGGTTTGGCTTCCACACGGAAACA-----TGT 814
QY 50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
Db 813 AGCCATGGGTGCTCTCTCCAGAGAGGTAGAGCGCTCTCTTATGGCCTCCAGTTTGTGA 754
QY 70 AspTyrSerIle----- 73
Db 753 GTTTTGTGCTAAGCTTTTCAGCATCTCACTCGCTGCAAAACGGTGCAGATGTGTCACCGG 694
QY 74 -----GluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHis-LeuThrMe 90
Db 693 CCAACCCGCGAACCCAGCCATCAT-----CATTTGACCCCT 658
QY 90 tlyGlySerSerMetLysCysSerSerLeu-----SerSerGluAlaIleLeuPh 107
Db 657 CCATTTCCCGAGTACCTCGTCTCTCTACTCTCTGACTCTCTACATGCTATCATGTT 598
QY 107 eThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCys-----LeuLe 124
Db 597 CACTATG-----TCTCTTGCTACACAGTGTGTGTTCTGCGCTGATTTCT 556
QY 124 uTyrLeuSer 127
Db 555 CTGCTTATCA 546
```

RESULT 11
US-09-186-188B-53/c
/ Sequence 53, Application US/09186188B
/ Patent No. 6455672
/ GENERAL INFORMATION:
/ APPLICANT: Benfey et al.
/ TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
/ FILE REFERENCE: 5914-074-999
/ CURRENT APPLICATION NUMBER: US/09/186,188B
/ PRIOR FILING DATE: 1998-11-05
/ PRIOR APPLICATION NUMBER: 08/842,445
/ PRIOR FILING DATE: 1997-04-24
/ PRIOR APPLICATION NUMBER: 08/638,617
/ PRIOR FILING DATE: 1996-04-26
/ NUMBER OF SEQ ID NOS: 79
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 53
/ LENGTH: 1093
/ TYPE: DNA
/ ORGANISM: Plant
US-09-186-188B-53

Alignment Scores:

Pred. No.: 0.5 Length: 1093
Score: 80.50 Matches: 37
Percent Similarity: 36.81% Conservative: 16
Best Local Similarity: 25.69% Mismatches: 47
Query Match: 11.60% Indels: 44
DB: 3 Gaps: 6

US-09-974-546C-84 (1-135) x US-09-186-188B-53 (1-1093)

```
QY 10 TyrGluAspMetHisAsnIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
Db 921 TACAGCGACAGATGTGTTTGTGCTCTCTTTATCCACTCAAGTAACCATCATCAC 862
QY 30 SerAenPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
Db 861 TATACTTACCCAAATATAGTTTGGGTTTGGCTTCCACACGGAAACA-----TGT 814
QY 50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
Db 813 AGCCATGGGTGCTCTCTCCAGAGAGGTAGAGCGCTCTCTTATGGCCTCCAGTTTGTGA 754
QY 70 AspTyrSerIle----- 73
Db 753 GTTTTGTGCTAAGCTTTTCAGCATCTCACTCGCTGCAAAACGGTGCAGATGTGTCACCGG 694
QY 74 -----GluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHis-LeuThrMe 90
Db 693 CCAACCCGCGAACCCAGCCATCAT-----CATTTGACCCCT 658
QY 90 tlyGlySerSerMetLysCysSerSerLeu-----SerSerGluAlaIleLeuPh 107
Db 657 CCATTTCCCGAGTACCTCGTCTCTCTACTCTCTGACTCTCTACATGCTATCATGTT 598
QY 107 eThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCys-----LeuLe 124
Db 597 CACTATG-----TCTCTTGCTACACAGTGTGTGTTCTGCGCTGATTTCT 556
QY 124 uTyrLeuSer 127
Db 555 CTGCTTATCA 546
```

RESULT 12
US-09-265-585C-53/c
/ Sequence 53, Application US/09265585C
/ Patent No. 6809234
/ GENERAL INFORMATION:
/ APPLICANT: Benfey, Phillip N.
/ APPLICANT: Di Laurenzio, Laura
/ APPLICANT: Wysocka-Diller, Joanna
/ APPLICANT: Malamy, Jocelyn E.
/ APPLICANT: Pysh, Leonard
/ APPLICANT: Helariutta, Yrjo
/ APPLICANT: Bruce, Wesley
/ APPLICANT: Lim, Jun
/ TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
/ FILE REFERENCE: 5914-066
/ CURRENT APPLICATION NUMBER: US/09/265,585C
/ CURRENT FILING DATE: 1999-03-10
/ PRIOR APPLICATION NUMBER: 08/842,445
/ PRIOR FILING DATE: 1997-04-24
/ PRIOR APPLICATION NUMBER: 08/638,617
/ PRIOR FILING DATE: 1996-04-26
/ NUMBER OF SEQ ID NOS: 152
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 53
/ LENGTH: 1093
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-09-265-585C-53

Alignment Scores:
Pred. No.: 0.5 Length: 1093
Score: 80.50 Matches: 37


```
Percent Similarity: 36.81% Conservative: 16
Best Local Similarity: 25.69% Mismatches: 47
Query Match: 11.60% Indels: 44
DB: 4 Gaps: 6

US-09-974-546C-84 (1-135) x US-09-265-585C-53 (1-1093)

QY 10 TyrGluaspMethHisAsnIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
Db 921 TACAGGACAGATGTGTTTGTGTGCTCTCTTTATCCACTCAAGTAACCATCATCAC 862
QY 30 SerAspPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
Db 861 TATACTTACCAATAGTTGGTTGGTTGGCTTCCACCGGAACA-----TGT 814
QY 50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69
Db 813 AGCCATGGGTGCTCTCTCCAGAAGAGGTAGACGCTCTCTTATGCGCTCCAGTTTGA 754
QY 70 AspTyrSerIle----- 73
Db 753 GTTTTGTCTAATAGTTTTCAGCATCTCTACCTGCTGCAACGCTGCAGATGTGCTGACCG 694
QY 74 -----GluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMe 90
Db 693 CCAACCGGTGAACCCAGCCATCAT-----CATTCTGACCTT 658
QY 90 tlysglyserSerMetLysCysSerSerLeu-----SerSerGluAlaIleLeuPh 107
Db 657 CCATTTCCTCCAGTACCTGCTCTCTCTACTCTCTGACTCTCTCATGCTCATCATGTT 598
QY 107 eThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCys-----LeuLe 124
Db 597 CACTATG-----TCTCTGTCTACACAGATGTGTCTGCGCTGATCTCT 556
QY 124 uTyrLeuSer 127
Db 555 CTGCTTATCA 546

RESULT 13
US-09-949-016-15729/c
; Sequence 99, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15729
; LENGTH: 148783
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(148783)
; OTHER INFORMATION: n = A,T,C or G

US-09-949-016-15729

Alignment Scores:
Pred. No.: 1.37e+03 Length: 148783
Score: 80.00 Matches: 41
Percent Similarity: 41.48% Conservative: 15
Best Local Similarity: 30.37% Mismatches: 52

Query Match: 11.53% Indels: 29
DB: 4 Gaps: 4

US-09-974-546C-84 (1-135) x US-09-949-016-15729 (1-148783)

QY 5 LeuArgAsnGlnLysTyrGluaspMethHisAsnIleHisIleLeuGlnIleArgLys 24
Db 15651 CTTTCATGCAATCCATATAAAGATACCACTATTATTATCCACATTTTACAGATAAGAAAA 15592
QY 25 LeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThrVal 44
Db 15591 CTGAGGCACAGAGA-CTAAATAAATCTAAATTTCCACAGCTAGTAAGT----- 15545
QY 45 LeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGlnLys 64
Db 15544 -----GGCAGAGCCCAAGAAATTAAGGTCAG 15521
QY 65 AlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIleLeuThrProLeu 84
Db 15520 GCCACTCTGATTC-----TCCAGCTTCTACCCAGTACACTATACCGTGTTCCTCATG-CTT 15468
QY 85 GlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSer----- 101
Db 15467 TCAACGCATTTCCCAATGGAGCAGGGGAGATAAGCGCAGCAACCAAGAGGACAT 15408
QY 102 -----SerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeu 117
Db 15407 CGACGCCAAAGGTCTACCATAAAGTTTTCACGGTGACATCCTTGCTT-----ACCTGT 15354
QY 118 GlyLeuGluCysCysLeuLeuTyrLysSerLysThrIleHisPro 132
Db 15353 CCAGGGCCATGCTGTGTGCATCTCTCTCCAAACCCGGGACCC 15309

RESULT 14
US-09-573-080A-99/c
; Sequence 99, Application US/09573080A
; Patent No. 6828097
; GENERAL INFORMATION:
; APPLICANT: JOAN, KNOLL
; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATION
; FILE REFERENCE: 30307
; CURRENT APPLICATION NUMBER: US/09/573,080A
; NUMBER OF SEQ ID NOS: 479
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 99
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: (1)-(262)
; OTHER INFORMATION: mir
; PUBLICATION INFORMATION:
; TITLE: Jurka, J; Walichiewicz, J; Milosavljevic, A
; JOURNAL: Journal of Molecular Evolution
; VOLUME: 35
; ISSUE: 4
; PAGES: 286-291
; DATE: 1992-10-
; DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
; DATABASE ENTRY DATE: 1996-01-26
; US-09-573-080A-99

Alignment Scores:
Pred. No.: 0.0721 Length: 262
Score: 79.50 Matches: 27
Percent Similarity: 46.67% Conservative: 1
Best Local Similarity: 45.00% Mismatches: 10
```

```
Query Match: 11.46% Indels: 22
DB: 4 Gaps: 2
US-09-974-546C-84 (1-135) x US-09-573-080A-99 (1-262)
QY 8 GlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHis 27
Db 174 CAACCCATGAGGTAGTACTATTATTATCCCATTTTACAGATGAGGAAACTGAGGCAC 115
QY 28 ArgLeu-----Ser 30
Db 114 AGAGAGTTAAGTAACTGCCAAGGTCACACAGTAGTAAGTGGCAGAGCCGGGATTGG 55
QY 31 AsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeu-ProPheCys 49
Db 54 AAC-----CCAGGCAGTCTGGCTCCAGAGYCCGTGCTCTTAAACCACTATGC 9
RESULT 15
US-08-675-566-5
; Sequence 5, Application US/08675566
; Patent No. 6090393
; GENERAL INFORMATION:
; APPLICANT: Fischer, Laurent
; TITLE OF INVENTION: PROMOTERS, EXPRESSION CASSETTES,
; TITLE OF INVENTION: RECOMBINANT VIRUSES, METHODS FOR MAKING, AND USES THEREOF
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/675,566
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2890
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)840-3333
; TELEFAX: (212)840-0712
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6196 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-675-566-5
Alignment Scores:
Pred. No.: 10.6 Length: 6196
Score: 79.50 Matches: 30
Percent Similarity: 46.40% Conservative: 28
Best Local Similarity: 24.00% Mismatches: 48
Query Match: 11.46% Indels: 19
DB: 3 Gaps: 4
US-09-974-546C-84 (1-135) x US-08-675-566-5 (1-6196)
QY 10 TyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
Db 1747 TACAAGACCTGATCCCTCCCTTAGC-----AGACAGCTCTTGCACACACGC 1794
QY 30 SerAsnPhePro-ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49
```

```
Db 1795 AGTAACTATCCACCGCTTTAAGAAAAAGCTTTAAGGCCAGCGCACATAACAGCTCCAATG-- 1852
QY 49 sTyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIl 69
Db 1853 -----TTTTTATCCAAGGAGAGCAAAATTTTCAGCAAGCGCAGGCTCAACAGTAATAGT 1905
QY 69 e-AspTyrSerIleGluGlnSerHisHisAlaIleLeuThrProLeuGlnThrHisLeuT 89
Db 1906 GAAGCAGAGGCATTTTCAGACGAGGCTCACTAGCTGCAGCTCGCCATTTATGAG----- 1958
QY 89 hrMetLysGlySerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrL 109
Db 1959 -----TCTGCAATAAAAAACAACACTCATCAGCAGCTGAAAAAGTGCACTTT---- 2003
QY 109 euThrLeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSerLysT 129
Db 2004 -----GACCTCATTAAGCCACTGCATATGCAAGTCTCTATATGCCGAGCCCGACA 2055
QY 129 hrIleHisPro 132
Db 2056 CCTCAATCCA 2066
```

Search completed: September 9, 2005, 17:01:29
Job time : 139.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 14:46:00 ; Search time 408.5 Seconds
(without alignments)
2170.324 Million cell updates/sec

Title: US-09-974-546C-84

Perfect score: 694

Sequence: 1 MRAPLRNQYEDMNIHL.....TLGLECCLLYLSKTHPQII 135

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09974546/runat_07092005_174504_21077/app.query.fasta_1.654
-DB=Published Applications NA -QWTF=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -MSAP=US09974546 @CGN 1.1.456 @runat_07092005_174504_21077
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOC=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubna/US05_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubna/US09_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/2/pubna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/2/pubna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/2/pubna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/2/pubna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/2/pubna/US10J_PUBCOMB.seq:
23: /cgn2_6/ptodata/2/pubna/US11A_PUBCOMB.seq:
24: /cgn2_6/ptodata/2/pubna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	694	100.0	2087	10	US-09-974-546-83
2	694	100.0	2505	10	US-09-974-546-85
3	90	13.0	1113	17	US-10-369-493-44085
4	90	13.0	1146	17	US-10-369-493-25021
5	87	12.5	1258	18	US-10-425-114-318
6	87	12.5	1624	20	US-10-425-115-49918
7	83	12.0	146547	15	US-10-017-128-1
8	82	11.8	24601	9	US-09-070-927A-223
9	81.5	11.7	133462	19	US-10-367-094-114
10	81	11.7	381	9	US-09-960-352-9738
11	81	11.7	819	19	US-10-437-963-23499
12	81	11.7	1277	20	US-10-425-115-137953
13	81	11.7	4437	20	US-10-425-115-178334
14	80.5	11.6	455	20	US-10-425-115-63753
15	80.5	11.6	600	22	US-10-972-079-4637
16	80.5	11.6	600	22	US-10-972-079-4638
17	80.5	11.6	1026	18	US-10-425-114-13022
18	80.5	11.6	1093	9	US-09-186-276B-53
19	80.5	11.6	1093	9	US-09-186-188B-53
20	80.5	11.6	1093	14	US-10-253-007-53
21	80.5	11.6	2343	20	US-10-425-115-127057
22	80	11.5	681	19	US-10-437-963-98199
23	80	11.5	1355	19	US-10-437-963-17754
24	80	11.5	1693	20	US-10-723-860-7296
25	79.5	11.5	262	10	US-09-854-867-99
26	79.5	11.5	262	21	US-10-786-970A-99
27	79.5	11.5	567	18	US-10-424-599-74400
28	79.5	11.5	585	20	US-10-357-930-54224
29	79	11.4	590	13	US-10-027-632-269701
30	79	11.4	590	17	US-10-027-632-269701
31	79	11.4	831	13	US-10-027-632-133030
32	79	11.4	831	17	US-10-027-632-133030
33	79	11.4	4068	19	US-10-437-963-35166
34	79	11.4	4107	20	US-10-425-115-139445
35	79	11.4	4335	20	US-10-425-115-139432
36	78.5	11.3	386	9	US-09-960-352-10443
37	78.5	11.3	600	22	US-10-972-079-88625
38	78.5	11.3	1787	13	US-10-027-632-84669
39	78.5	11.3	1787	13	US-10-027-632-84670
40	78.5	11.3	1787	17	US-10-027-632-84669
41	78.5	11.3	1787	17	US-10-027-632-84670
42	78	11.2	650	13	US-10-027-632-223621
43	78	11.2	650	17	US-10-027-632-223621
44	78	11.2	1207	18	US-10-424-599-5755
45	78	11.2	9869	20	US-10-425-115-139430

ALIGNMENTS

RESULT 1

US-09-974-546-83
; Sequence 83, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert

TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS, PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE

NUMBER OF SEQUENCES: 87

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/974,546
FILING DATE: 10-Oct-2001
CLASSIFICATION: Unknown
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/097,199
FILING DATE: 1998-06-12
ATTORNEY/AGENT INFORMATION:
NAME: Nakashima, Richard A.
REGISTRATION NUMBER: P-42,023
REFERENCE/DOCKET NUMBER: UROC:018
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 2087 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 99..503
SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-974-546-83

Alignment Scores:
Pred. No.: 5,82e-85 Length: 2087
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-974-546-83 (1-2087)

QY 1 MetArgAlaPheLeuArgAenGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCCTTTAAGGAAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAenPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAATTTGAGGCACAGATTAAGTAATCTCCCAAGGCTACAGGCATCTTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60
Db 219 CCAGAAACTGTGCTCTTACCATCTGCTCAAGGATTTTCGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 279 AGAAGTCAAAAGCAACAGATTCATTGATTATTCATAGACAGTCACACCATGCATTT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
Db 339 CTCACACCCCTTGACACACATTTGACATGAAAGGTTCTCAATGAAATGTTCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 399 TCTTCAGAGCCATATATTACATTTGACTTTGACGTTAACTCAACCCCTAGGCTCGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 459 TGCTGTCTTCTACTATATCCAAACATATACATCCACAGATCATATA 503

RESULT 2

US-09-974-546-85
Sequence 85, Application US/09974546
Publication No. US20030050470A1
GENERAL INFORMATION:
APPLICANT: An, Gang
O'Hara, S. Mark
Ralph, David
Veltri, Robert
TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/974,546
FILING DATE: 10-Oct-2001
CLASSIFICATION: Unknown
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/097,199
FILING DATE: 1998-06-12
ATTORNEY/AGENT INFORMATION:
NAME: Nakashima, Richard A.
REGISTRATION NUMBER: P-42,023
REFERENCE/DOCKET NUMBER: UROC:018
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 2505 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 99..503
SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-974-546-85

Alignment Scores:
Pred. No.: 7,74e-85 Length: 2505
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-974-546C-84 (1-135) x US-09-974-546-85 (1-2505)

QY 1 MetArgAlaPheLeuArgAenGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCCTTTAAGGAAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAenPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAATTTGAGGCACAGATTAAGTAATCTCCCAAGGCTACAGGCATCTTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60
Db 219 CCAGAAACTGTGCTCTTACCATCTGCTCAAGGATTTTCGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: September 9, 2005, 13:09:35 ; Search time 1867.5 Seconds
(without alignments)
2751.634 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLNQKEDMNIHL.....TLGLECLLYLSKTHPQII 135

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-O=/cgn2.1/USPTO.spool/US9974546/runat.07092005.174502.20989/app.query.fasta_1.654
-DB=EST -CPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US9974546 @CNG 1.1 3556 @runat.07092005.174502.20989 -NCPUP=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gsel:
9: gb_gsel2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	97	14.0	690	5	BU242169
2	95.5	13.8	588	5	BW569427
3	94	13.5	606	6	CB286029
4	90.5	13.0	459	5	BW557636
5	90.5	13.0	543	5	BW582739
6	90	13.0	570	5	BW591794
7	89.5	12.9	1151	5	BU413225
8	89	12.8	811	5	BU436282
9	88.5	12.8	743	7	CV511381
					CV511381 kc40h08.y

ALIGNMENTS

RESULT 1
BU242169/c
LOCUS BU242169 CSEQCHN34 Gallus gallus linear EST 26-NOV-2002
DEFINITION 603779592F1 CSEQCHN34 Gallus gallus cDNA clone ChEST725h23 5', mRNA
sequence.
ACCESSION BU242169
VERSION BU242169.1 GI:25488447
KEYWORDS EST,
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 690)
AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. .690
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton line 151"

10 88.5 12.8 789 8 BH559290
11 88 12.7 776 9 AG377362
12 87.5 12.6 803 5 BX756993
13 87 12.5 788 9 CR262699
14 86.5 12.5 741 3 AY439721
15 86 12.4 529 5 BO592283
16 86 12.4 769 8 BZ521554
17 86 12.4 992 9 CL254017
18 85.5 12.3 788 8 CC091268
19 85.5 12.3 877 6 CA472302
20 85 12.2 579 9 CE464906
21 85 12.2 707 9 CC743728
22 85 12.2 906 9 CG300818
23 85 12.2 914 9 CG300807
24 84.5 12.2 285 4 BM111157
25 84.5 12.2 562 8 AZ954412
26 84.5 12.2 652 2 BF644675
27 84.5 12.2 688 5 BP027905
28 84.5 12.2 741 4 BI309709
29 84.5 12.2 770 4 BM780176
30 84 12.1 493 5 BX836326
31 84 12.1 624 8 BZ745987
32 84 12.1 728 6 CD821188
33 84 12.1 896 9 CG008627
34 84 12.1 916 7 CK151729
35 84 12.1 969 2 BE739923
36 84 12.1 1399 2 AW730867
37 83.5 12.0 721 8 B69065
38 83.5 12.0 732 5 BU264381
39 83.5 12.0 781 7 CN135272
40 83.5 12.0 952 1 AL665968
41 83.5 12.0 1335 4 BM467342
42 83 12.0 395 6 CB069753
43 83 12.0 572 2 BE015616
44 83 12.0 612 2 BE015598
45 83 12.0 634 9 AG110445

BH559290 BOHOW80TR
AG377362 Mus muscu
BX756993 BX756993
CR262699 Reverse s
AY439721 Armigeres
BO592283 E012698-0
BZ521554 BOKAB11PF
CL254017 ZMMBB060
CC091268 CSU-K33r.
CA472302 AGENCOURT
CE464906 tigr-gss-
CC743728 ZMMBB011
CG300818 OG0AB25TV
CG300807 OG0AB25TH
BM111157 EST558693
AZ954412 ZM022005
BF644675 NF019F01E
BP027905 BP027905
BI309709 EST531119
BM780176 EST590764
BX836326 BX836326
BZ745987 OGF8K05TM
CD821188 BN25.041B
CG008627 ZUAC085TH
CK151729 FGAS03441
BE739923 601556175
AW730867 GA_Ea002
B69065 CIT-HSP-205
BU264381 603505585
CN135272 OX1_31_E0
AL665968 AL665968
BM467342 AGENCOURT
CB069753 la28d02.x
BE015616 EST842 Ma
BE015598 EST824 Ma
AG110445 Pan trogl

/db_xref="taxon:9031"
 /clone="ChEST725h23"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSBQCN34"

/note="Organ: liver; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylation C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was bluntended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Alignment Scores:
 Pred. No.: 1.02 Length: 690
 Score: 97.00 Matches: 33
 Percent Similarity: 45.08% Conservative: 22
 Best Local Similarity: 27.05% Mismatches: 59
 Query Match: 13.98% Indels: 8
 DB: 5 Gaps: 3

US-09-974-546C-84 (1-135) x BU242169 (1-690)

QY 15 AaIIeIIeHISleuGlnleuArgLysLeuArgHisArgLeuSerAsnPhProArg 34
 Db 623 AACTTAATAGCATTTTACGTTTCCCTACAGTTACTAAAGAAATATCAAT-----GAT 570
 QY 35 LeuProGlyIleLeuAlaPro-----GluThrValLeuLeuProPheCysTyrLys 51
 Db 569 GTCCAGGACTCTGTGTCACAGAGGAGACAGATTTGCTCAGCATATCAGATG 510
 QY 52 ValPheArgLysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyr 71
 Db 509 AAATTTAGGTGAAATAAAGACAGTATACACTGATTAAGAGCTGCAGAG-----AAT 459
 QY 72 SerIleGluGlnSerHisAlaIleLeuThrProLeuGluThrHisLeuThrMetLys 91
 Db 458 TCAGTTATAAAGGCCATTTCTATTTCAGAGCTTTGCTAGGATGCCTCTCTTATAAGA 399
 QY 92 GlySerSerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
 Db 398 AAAAAAGGAAAAAGAAAGACACCAACCACTGAGATGCTGCTATTCTCACTAAATGC 339
 QY 112 GlnLeuThrGlnThrLeuGlyLeuGluCysLeuLeuTyrLeuSerLysThrIleHis 131
 Db 338 AAGGCCACCACTACATCTCTGTTTCTGCTGCTCTGTTTACAGGAGTAACCATTCAT 279
 QY 132 ProGln 133
 Db 278 CCAAG 273

RESULT 2

BW569427 588 bp mRNA linear EST 31-AUG-2004
 LOCUS BW569427 Yutaka Satou unpublished cDNA library (cstb) Ciona
 DEFINITION savignyi cDNA clone cstb014a08 3', mRNA sequence.

ACCESSION BW569427
 VERSION BW569427
 KEYWORDS EST.
 SOURCE Ciona savignyi
 ORGANISM Ciona savignyi

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Ciona.
 REFERENCE 1 (bases 1 to 588)

AUTHORS Satou,Y. and Satoh,N.
 TITLE Expressed genes in Ciona savignyi
 JOURNAL Unpublished (2004)
 COMMENT Contact: Yutaka Satou
 Department of Zoology
 Kyoto University
 Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4095
 Fax: 81-75-705-1113
 Email: yutaka@ascidian.zool.kyoto-u.ac.jp.

FEATURES

source
 1..588
 Location/Qualifiers
 /organism="Ciona savignyi"
 /mol_type="mRNA"
 /db_xref="taxon:51511"
 /clone="cstb014a08"
 /dev_stage="tailbud stage"
 /clone_lib="Yutaka Satou unpublished cDNA library (cstb)"

ORIGIN

Alignment Scores:
 Pred. No.: 1.23 Length: 588
 Score: 95.50 Matches: 37
 Percent Similarity: 44.27% Conservative: 21
 Best Local Similarity: 28.24% Mismatches: 48
 Query Match: 13.76% Indels: 25
 DB: 5 Gaps: 6

US-09-974-546C-84 (1-135) x BW569427 (1-588)

QY 24 LysLeuArgHisArgLeuSerAsn-----PheProArgLeuProGlyIleLeu 39
 Db 82 AAATATTACATAAATACTTGGCAACAAAACACAGATTTTACAACTTGCCTAGTTGA 141
 QY 40 AlaProGluThrValLeuLeuProPheCysTyrLysValPhe-----ArgLys 55
 Db 142 CTAGAAGAATCCATTTTTTAAAGAGCAAGCAAAATTTACATAATTTGCAATAATCCG 201
 QY 56 LysGluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGln 75
 Db 202 TAGGAATTCCTTAGAAGATTTTAAAGAGAGAGCGATTTTCGTAGATATAGCAGCAACAA 261
 QY 76 Ser-----HisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysLysSer 93
 Db 262 AGCAATTTGCACACAGCTATATTTGCAGGGATTCAGTCATTTCTTCCACTGAGTGCCTCA 321
 QY 94 SerMetLysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeu 113
 Db 322 AGAGAGCACTGTTGCTTGTGCGAGCTGCTGATCTGCCTC-----CAGCGGATT 369
 QY 114 ThrGlnThrLeuGlyLeuGluCys-----CysLeuLeu----- 124
 Db 370 GGGAGCTGGGTGGGTCTGTTGTCATTCGTCAAGAAATTTGAGGGTGTCTGTTGTTGCCAC 429
 QY 125 -----TyrLeuSerLysThrIleHisProGln 133
 Db 430 TACTCTACCTTGTGATGCTTTTCTCCACAA 462

RESULT 3

CB286029 606 bp mRNA linear EST 27-FEB-2003
 LOCUS CB286029
 DEFINITION CMD29.B11_90 UMNPM3 Sus scrofa cDNA clone PPSUBLB_15B11 5', mRNA sequence.

ACCESSION CB286029
 VERSION CB286029.1
 KEYWORDS GI:28576983
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 606)
 AUTHORS Dvorak,C.M.T., Hyland,K.A., Zhang,Y., Fahrenkrug,S.C. and Murtaugh,M.P.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:23:48 ; Search time 58 Seconds
(without alignments)
900.218 Million cell updates/sec

Title: US-09-974-546C-84
Perfect score: 694
Sequence: 1 MRAFLRNQYEDWHNIHL.....TLGLSCLLYLSKTHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Genesepc16Dec04:*
1: Genesepc19808:*
2: Genesepc19908:*
3: Genesepc20008:*
4: Genesepc20018:*
5: Genesepc20028:*
6: Genesepc20038a:*
7: Genesepc20038b:*
8: Genesepc20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	694	100.0	135	3 AAY59296	Prostate
2	694	100.0	135	3 AAY59295	Prostate
3	694	100.0	135	4 AAU02174	Biomarker
4	694	100.0	135	4 AAU02175	Biomarker
5	103	14.8	21	3 AAY59294	Cancer bi
6	103	14.8	21	4 AAU02173	Biomarker
7	90	13.0	371	8 ADS41968	Bacterial
8	90	13.0	381	8 ADN18681	Bacterial
9	82.5	11.9	645	5 ABB90843	Herbicida
10	81	11.7	384	8 ADN48099	Thermococ
11	77.5	11.2	1061	7 ADF50279	Human PFM
12	77.5	11.2	1061	8 ADO36741	Human PR-
13	77.5	11.2	1144	4 ABB58124	Drosophil
14	77	11.1	1026	8 ADS08184	Staphyloc
15	76.5	11.0	322	6 ABM69043	Phototrab
16	75	10.8	633	6 ABU42279	Protein e
17	73	10.5	195	4 ABG18410	Novel hum
18	72	10.4	523	4 AAB46398	H. pylori
19	72	10.4	636	8 ADM94232	Soybean E
20	71.5	10.3	450	7 ADF41696	Bacillus
21	71.5	10.3	634	8 ADM94288	Corn Ethy
22	71.5	10.3	676	4 ABG15498	Novel hum
23	71	10.2	127	4 AAU42328	Propionib
24	71	10.2	127	6 ABM38847	Propionib
25	71	10.2	424	4 AAM39676	Human pol

26	71	10.2	424	4 AAG81358	Human APP
27	71	10.2	424	8 ADH09599	Human hos
28	71	10.2	424	8 ADH09598	Human hos
29	71	10.2	424	8 ADH09600	Human hos
30	71	10.2	429	4 AAM41462	Human pol
31	71	10.2	519	4 AAB92584	Human pro
32	71	10.2	1922	8 ADL97802	Human dcr
33	71	10.2	1924	4 AAE09768	Human dlc
34	71	10.2	1924	5 AAO15987	Human hel
35	71	10.2	1924	6 ABO07204	Human p53
36	71	10.2	1924	7 ABU64479	Stem loop
37	71	10.2	1924	7 ADN60159	Human hel
38	70.5	10.2	438	7 ABO64633	Klebeieil
39	70.5	10.2	634	8 ADM94300	Corn Ethy
40	70	10.1	206	7 ADJ69892	Human hea
41	70	10.1	982	8 ADS29846	Bacterial
42	69.5	10.0	1294	2 AAW30601	Human typ
43	69.5	10.0	1353	8 ADO89904	Antagonis
44	69.5	10.0	1398	3 AAB18292	Plasmodiu
45	69.5	10.0	3105	5 ABB80604	Human sbg

ALIGNMENTS

RESULT 1
AAY59296
ID AAY59296 standard; peptide; 135 AA.
XX AC AAY59296;
XX AC
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 amino acid sequence.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
XX
PR N-PSDB; AAZ87584.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
XX breast and bladder cancer.
XX
PS Example 5; Page 184-186; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer.
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences from various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used

CC to detect antibodies. The proteins and antibodies can be used in
 CC immunodetection methods for detecting or quantifying the cancers, and for
 CC clinical diagnosis of these cancers. The antibodies may also be used for
 CC radioimaging to quantify and localize the encoded proteins

XX Sequence 135 AA;
 SQ
 Query Match 100.0%; Score 694; DB 3; Length 135;
 Best Local Similarity 100.0%; Pred. No. 6.9e-75;
 Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRAFLRNQKYEDMHNIIHILOIRKLHRLSNPRLPGILAPETVLLPFCYKVRKKEKVK 60
 DB 1 MRAFLRNQKYEDMHNIIHILOIRKLHRLSNPRLPGILAPETVLLPFCYKVRKKEKVK 60
 QY 61 RSQKATEFDYISIEQSHHAILTPLQHLTMKGSSMKCSSLSSEAILFTLTQLTQTGLG 120
 DB 61 RSQKATEFDYISIEQSHHAILTPLQHLTMKGSSMKCSSLSSEAILFTLTQLTQTGLG 120
 QY 121 CCLLYLSKTIHPQII 135
 DB 121 CCLLYLSKTIHPQII 135

RESULT 2
 AAY59295
 ID AAY59295 standard; peptide; 135 AA.
 AC AAY59295;
 XX
 DT 19-APR-2000 (first entry)
 DE Prostate disease marker UC Band #28 amino acid sequence.
 KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
 KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
 OS Homo sapiens.
 XX
 PN WO9964631-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US013151.
 XX
 PR 12-JUN-1998; 98US-00097199.
 XX
 PA (UROC-) UROCOR INC.
 XX
 PI An G, O'hara SM, Ralph D, Veltri RW;
 XX
 DR WPI; 2000-116557/10.
 DR N-PSDB; AA287583.
 XX

Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
 PT breast and bladder cancer.
 XX
 PS Example 5; Page 182-183; 191pp; English.

XX The invention provides nucleic acid markers of prostate, breast and
 CC bladder cancer. The markers are indicators of malignant transformation of
 CC prostate, breast and bladder tissues and are diagnostic of the potential
 CC for metastatic spread of malignant prostate tumours. The nucleic acid can
 CC also be used as targets for therapeutic intervention in prostate cancer,
 CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
 CC markers may be used to design specific probes and primers, for the rapid
 CC analysis of prostate, bladder or breast biopsy samples. The probes and
 CC primers may also be used for in situ hybridization or in situ PCR
 CC detection and diagnosis. They may also be used to identify and isolate
 CC full length gene sequences from various DNA libraries. Antibodies against
 CC the polypeptide products of the markers can be used to treat prostate
 CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
 CC to detect antibodies. The proteins and antibodies can be used in

CC immunodetection methods for detecting or quantifying the cancers, and for
 CC clinical diagnosis of these cancers. The antibodies may also be used for
 CC radioimaging to quantify and localize the encoded proteins

XX Sequence 135 AA;
 SQ
 Query Match 100.0%; Score 694; DB 3; Length 135;
 Best Local Similarity 100.0%; Pred. No. 6.9e-75;
 Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRAFLRNQKYEDMHNIIHILOIRKLHRLSNPRLPGILAPETVLLPFCYKVRKKEKVK 60
 DB 1 MRAFLRNQKYEDMHNIIHILOIRKLHRLSNPRLPGILAPETVLLPFCYKVRKKEKVK 60
 QY 61 RSQKATEFDYISIEQSHHAILTPLQHLTMKGSSMKCSSLSSEAILFTLTQLTQTGLG 120
 DB 61 RSQKATEFDYISIEQSHHAILTPLQHLTMKGSSMKCSSLSSEAILFTLTQLTQTGLG 120
 QY 121 CCLLYLSKTIHPQII 135
 DB 121 CCLLYLSKTIHPQII 135

RESULT 3
 AAU02174
 ID AAU02174 standard; protein; 135 AA.
 XX
 AC AAU02174;
 XX
 DT 29-AUG-2001 (first entry)
 DE Biomarker protein encoded by UC band 28 #2, used in diagnosis of cancer.
 KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
 KW benign prostatic hyperplasia; BPH; therapeutic; human.
 OS Homo sapiens.
 XX
 PN US6218529-B1.
 XX
 PD 17-APR-2001.
 XX
 PF 12-JUN-1998; 98US-00097199.
 XX
 PR 31-JUL-1995; 95US-0001655P.
 PR 11-JAN-1996; 96US-0013611P.
 PR 31-JUL-1996; 96US-00692787.
 XX
 PA (UROC-) UROCOR INC.
 XX
 PI An G, O'hara SM, Ralph D, Veltri R;
 XX
 DR WPI; 2001-289849/30.
 DR N-PSDB; AAS04000.
 XX

New nucleic acids as biomarkers and targets useful for detecting,
 PT diagnosing, prognosing, and in developing treatments for prostate, breast
 PT and bladder cancer.
 XX
 PS Disclosure; Col 121; 78pp; English.
 XX
 CC The sequence represents the amino acid sequence of biomarker protein
 CC encoded by UC band 28 #2. Proteins encoded by the nucleic acid markers
 CC can be used to produce antibodies for the detection of prostate, breast
 CC or bladder cancer. Biomarker nucleic acid sequences can be used as
 CC hybridisation probes and primers that specifically hybridise to prostate
 CC cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast
 CC cancer markers. The nucleic acids can be used as targets for therapeutic
 CC intervention in these diseases, in the identification and isolation of
 CC full-length gene sequences, including regulatory elements for gene
 CC expression, from genomic human DNA libraries, as hybridisation probes for
 CC screening genomic human DNA libraries. The kits comprising the nucleic
 CC acid sequences are useful for detecting bladder, breast or prostate


```
CC cancer cells in a biological sample
SQ Sequence 135 AA;

Query Match      100.0%; Score 694; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9e-75;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLRNQKVEDMHNIIHILOIKLRHLSNFPRLPGILAPETVLLPFCVKVFRKKEVK 60
   |||||
DB 1 MRAFLRNQKVEDMHNIIHILOIKLRHLSNFPRLPGILAPETVLLPFCVKVFRKKEVK 60
   |||||
QY 61 RSQKATEFIDYSIEQSHHAILTPLOTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
   |||||
DB 61 RSQKATEFIDYSIEQSHHAILTPLOTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
   |||||
QY 121 CCLLYLSKTIHPQII 135
   |||||
DB 121 CCLLYLSKTIHPQII 135
   |||||

RESULT 4
AAU02175
ID AAU02175 standard; protein; 135 AA.
AC AAU02175;
XX
XX
DT 29-AUG-2001 (first entry)
XX
XX Biomarker protein encoded by UC band 28 #3, used in diagnosis of cancer.
XX
XX Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
XX benign prostatic hyperplasia; BPH; therapeutic; human.
XX
XX Homo sapiens.
XX
XX US6218529-B1.
XX
XX 17-APR-2001.
XX
XX 12-JUN-1998; 98US-00097199.
XX
XX 31-JUL-1995; 95US-0001655P.
XX
XX 11-JAN-1996; 96US-0013611P.
XX
XX 31-JUL-1996; 96US-00692787.
XX
XX (UROC-) UROCOR INC.
XX
XX An G, O'hara SM, Ralph D, Veltri R;
XX
XX WPI; 2001-289849/30.
XX
XX N-PSDB; AAS04001.
XX
XX New nucleic acids as biomarkers and targets useful for detecting,
XX diagnosing, prognosing, and in developing treatments for prostate, breast
XX and bladder cancer.
XX
XX Disclosure; Col 125; 78pp; English.
XX
XX The sequence represents the amino acid sequence of biomarker protein
XX encoded by UC band 28 #3. Proteins encoded by the nucleic acid markers
XX can be used to produce antibodies for the detection of prostate, breast
XX or bladder cancer. Biomarker nucleic acid sequences can be used as
XX hybridisation probes and primers that specifically hybridise to prostate
XX cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast
XX cancer markers. The nucleic acids can be used as targets for therapeutic
XX intervention in these diseases, in the identification and isolation of
XX full-length gene sequences, including regulatory elements for gene
XX expression, from genomic human DNA libraries, as hybridisation probes for
XX screening genomic human DNA libraries. The kits comprising the nucleic
XX acid sequences are useful for detecting bladder, breast or prostate
XX cancer cells in a biological sample
XX
```

```
SQ Sequence 135 AA;

Query Match      100.0%; Score 694; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9e-75;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLRNQKVEDMHNIIHILOIKLRHLSNFPRLPGILAPETVLLPFCVKVFRKKEVK 60
   |||||
DB 1 MRAFLRNQKVEDMHNIIHILOIKLRHLSNFPRLPGILAPETVLLPFCVKVFRKKEVK 60
   |||||
QY 61 RSQKATEFIDYSIEQSHHAILTPLOTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
   |||||
DB 61 RSQKATEFIDYSIEQSHHAILTPLOTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
   |||||
QY 121 CCLLYLSKTIHPQII 135
   |||||
DB 121 CCLLYLSKTIHPQII 135
   |||||

RESULT 5
AA59294
ID AA59294 standard; peptide; 21 AA.
XX
XX AC AA59294;
XX
XX 19-APR-2000 (first entry)
XX
XX Cancer biomarker UC28 antigenic peptide.
XX
XX Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
XX benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
XX Homo sapiens.
XX
XX WO9964631-A1.
XX
XX 16-DEC-1999.
XX
XX 11-JUN-1999; 99WO-US013151.
XX
XX 12-JUN-1998; 98US-00097199.
XX
XX (UROC-) UROCOR INC.
XX
XX An G, O'hara SM, Ralph D, Veltri RW;
XX
XX WPI; 2000-116557/10.
XX
XX Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
XX breast and bladder cancer.
XX
XX Example 5; Page 114; 191pp; English.
XX
XX The invention provides nucleic acid markers of prostate, breast and
XX bladder cancer. The markers are indicators of malignant transformation of
XX prostate, breast and bladder tissues and are diagnostic of the potential
XX for metastatic spread of malignant prostate tumours. The nucleic acid can
XX also be used as targets for therapeutic intervention in prostate cancer.
XX benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
XX markers may be used to design specific probes and primers, for the rapid
XX analysis of prostate, bladder or breast biopsy samples. The probes and
XX primers may also be used for in situ hybridization or in situ PCR
XX detection and diagnosis. They may also be used to identify and isolate
XX full length gene sequences from various DNA libraries. Antibodies against
XX the polypeptide products of the markers can be used to treat prostate
XX cancer, bladder cancer or breast cancer. The encoded proteins may be used
XX to detect antibodies. The proteins and antibodies can be used in
XX immunodetection methods for detecting or quantifying the cancers, and for
XX clinical diagnosis of these cancers. The antibodies may also be used for
XX radioimaging to quantify and localize the encoded proteins
XX
XX Sequence 21 AA;
XX
```

```
Query Match      14.8%; Score 103; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 RKKEKVRKSKATEFTDYSIE 74
DB 1 RKKEKVRKSKATEFTDYSIE 21

RESULT 6
AAU02173
ID AAU02173 standard; peptide; 21 AA.
AC AAU02173;
XX AC
XX 29-AUG-2001 (first entry)
XX
XX Biomarker UC band 28, antigenic peptide used in diagnosis of cancer.
DE
XX
XX Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; antigenic.
KW
XX Homo sapiens.
OS
XX US6218529-B1.
XX
XX 17-APR-2001.
XX
XX 12-JUN-1998; 98US-00097199.
XX
XX 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.
XX
XX (UROC-) UROCOR INC.
PA
XX
XX An G, O'hara SM, Ralph D, Veltri R;
PI
XX WPI; 2001-289849/30.
XX
XX New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
XX Example 5; Col 73; 78pp; English.
XX
XX The sequence represents the amino acid sequence of biomarker, UC band 28,
CC antigenic peptide used to produce antibodies for the detection of
CC prostate, breast or bladder cancer. Biomarker nucleic acid sequences can
CC be used as hybridisation probes and primers that specifically hybridise
CC to prostate cancer, benign prostatic hyperplasia (BPH), bladder cancer or
CC breast cancer markers. The nucleic acids can be used as targets for
CC therapeutic intervention in these diseases, in the identification and
CC isolation of full-length gene sequences, including regulatory elements
CC for gene expression, from genomic human DNA libraries, as hybridisation
CC probes for screening genomic human DNA libraries. The kits comprising the
CC nucleic acid sequences are useful for detecting bladder, breast or
CC prostate cancer cells in a biological sample
XX
XX Sequence 21 AA;
SQ

Query Match      14.8%; Score 103; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 RKKEKVRKSKATEFTDYSIE 74
DB 1 RKKEKVRKSKATEFTDYSIE 21

RESULT 7
ADS41968
ID ADS41968 standard; protein; 371 AA.
XX
XX ADS41968;
XX
XX 02-DEC-2004 (first entry)
XX
XX Bacterial polypeptide #20398.
XX
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
XX Bacteria.
OS
XX US2003233675-A1.
XX
XX 18-DEC-2003.
XX
XX 20-FEB-2003; 2003US-00369493.
XX
XX 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 20398; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 371 AA;
SQ

Query Match      13.0%; Score 90; DB 8; Length 371;
Best Local Similarity 26.8%; Pred. No. 0.1;
Matches 34; Conservative 23; Mismatches 38; Indels 32; Gaps 7;

QY 13 MNHIIHILQIRKLHRL-----SNFP-----RLPGILAPETVLLPFCYKVF 53
```

Db 21 MHNL--AIKLRGHEVGIVTNNRPTGKBEELKRYGIELIKIPGIISP-----FLDV 70
Qy 54 RKKEKVRKSKATEFI-DYSIEQSHHAILTPLOTHLTMKGSSM-KCSSLSSEAILFTLTL 111
Db 71 NLTYGLKSSEELNEFLKDFDIHSHHA-FTPLSLKALKAGKQKMEKGTLLTTHSISPAHES 129
Qy 112 QLTQTILG 118
Db 130 KLWDTLG 136

RESULT 8
ADN18681
ID ADN18681 standard; protein; 381 AA.
XX AC
XX ADN18681;
XX 02-DEC-2004 (first entry)
XX Bacterial polypeptide #1334.
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX Bacteria.
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOV/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 1334; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or

CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 381 AA;
Query Match 13.0%; Score 90; DB 8; Length 381;
Best Local Similarity 26.8%; Pred. No. 0.11;
Matches 34; Conservative 23; Mismatches 38; Indels 32; Gaps 7;
Qy 13 MNNIHILOIRKLRLHRL-----SNFP-----RLPGLIAPETVLPPCYKVF 53
Db 24 MHNL--AIKLRGHEVGIVTNNRPTGKBEELKRYGIELIKIPGIISP-----FLDV 73
Qy 54 RKKEKVRKSKATEFI-DYSIEQSHHAILTPLOTHLTMKGSSM-KCSSLSSEAILFTLTL 111
Db 74 NLTYGLKSSEELNEFLKDFDIHSHHA-FTPLSLKALKAGKQKMEKGTLLTTHSISPAHES 132
Qy 112 QLTQTILG 118
Db 133 KLWDTLG 139

RESULT 9
ABB90843
ID ABB90843 standard; protein; 645 AA.
XX AC ABB90843;
XX 31-MAY-2002 (first entry)
XX Herbicidally active polypeptide SEQ ID NO 54.
XX Herbicidal; plant; agriculture; herbicide.
XX Arabidopsis thaliana.
XX WO200210210-A2.
XX 07-FEB-2002.
XX 28-AUG-2001; 2001WO-EP009892.
XX 28-AUG-2001; 2001WO-EP009892.
XX (FARB) BAYER AG.
XX Tietjen K, Weidler M;
XX WPI; 2002-269010/31.
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.
XX Claim 5; SEQ ID NO 54; 261pp + Sequence Listing; English.
XX The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX Sequence 645 AA;
SQ

Query Match 11.9%; Score 82.5; DB 5; Length 645;
 Best Local Similarity 23.5%; Pred. No. 1.7;
 Matches 31; Conservative 26; Mismatches 46; Indels 29; Gaps 5;

QY 25 LHRHLSNP---PRLPGILAPETV-----LLPFCYKVRKKVKRSQ 63
 DB 99 MTHLLAGTYGHPWVMTAVTFKMLTGIVSPLTALSIVTLPLLLKA--KVREPMLSK 156

QY 64 KATEFDYIEQSHHAILPLOTHTLTKGSSMKC--SSLSSEAILFTLTQLTQTGLECC 122
 DB 157 KTRF-----LDREVGLIMQKTESLHVRMLTTKIRTSLDRTHTLYTLVELSKTLGLKNC 211

QY 123 LLYLSKTIHPQI 134
 DB 212 AVWIPNEIKTEM 223

RESULT 10
 ADN48099
 ID ADN48099 standard; protein; 384 AA.
 AC ADN48099;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Thermococcus kodakaraensis KOD1 protein sequence SeqID1977.
 XX
 KW gene disruption; gene targeting; marker gene; transformation;
 KW homologous recombination; hyperthermostable archaeobacterium; KOD1;
 KW gene structure; gene function; enzyme activity; medicine;
 KW forensic science; food; drug inspection; molecular biology; immunology.
 XX
 OS Thermococcus kodakaraensis.
 XX
 FN WO2004022736-A1.
 XX
 PD 18-MAR-2004.
 XX
 PF 29-AUG-2003; 2003WO-IB003597.
 XX
 PR 30-AUG-2002; 2002JP-00319011.
 XX
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 XX
 PI Imanaka T, Atomi H;
 XX
 XX WPI; 2004-257593/24.
 XX
 PT Method for disrupting targeted gene in genome of organism particularly
 PT thermostable bacterium and with genome chips for analysis, applicable in
 PT studying gene structure and functions.
 XX
 PS Claim 9; SEQ ID NO 1977; 598pp; Japanese.
 XX
 CC This invention relates to a novel method for targeting disruption of an
 CC arbitrary gene in a genome of an organism which comprises providing the
 CC whole sequential data of the genome of such organism, selecting at least
 CC 1 arbitrary region in the sequence, providing a vector that contains a
 CC sequence homologous with the selected region and a marker gene,
 CC transformation, and homologous recombination. The genome is preferably
 CC the genome of a hyperthermostable archaeobacterium, particularly
 CC Thermococcus kodakaraensis KOD1. The method is for targeting the
 CC disruption of a gene in the genome of an organism, which is applicable in
 CC studying gene structure and functions as well as enzyme activities of
 CC encoded proteins and useful in medicine, forensic science, food or drug
 CC inspection, molecular biology and immunology. With this method, the
 CC disruption of a gene at an arbitrary position in a genome can be achieved
 CC efficiently and reliably. The present sequence is that of a protein
 CC encoded by the genome of Thermococcus kodakaraensis which was derived
 CC using the method of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 384 AA;
 Query Match 11.7%; Score 81; DB 8; Length 384;
 Best Local Similarity 23.3%; Pred. No. 1.3;
 Matches 31; Conservative 28; Mismatches 42; Indels 32; Gaps 7;

QY 13 MNIIHILQIRKLURLS-----NFPRLPGILAPETVLLPFCYKV 53
 DB 21 MQL--AYLKKGLGHDVSVTNDLTGKKELEGLVGVGVISP-VLGINITY--- 74

QY 54 RKKEKVKRSQKATEF-IDYSIEQSHHAILPLOTHTLTKGSSM-KCSSLSSEAILFTLT 111
 DB 75 ---GLKSNRELGEPLVDFDVVHAHA-FTPLSLKAVKAGRTLEKATLLTTHSISFSHES 129

QY 112 QLTQTGLECCLL 124
 DB 130 SLWKALGLTFPLL 142

RESULT 11
 ADF50279
 ID ADF50279 standard; protein; 1061 AA.
 XX
 AC ADF50279;
 XX
 DT 12-FEB-2004 (first entry)
 XX
 DE Human PPM7 protein, a PR/SET family member.
 XX
 KW human; PPM7; SET; Su(var)3-9, Enhancer-of-Zeste and Trithorax; cancer;
 KW PPM/SET; tumour suppressor; hyperproliferative disorder; scleroderma;
 KW arthritis; alcoholic liver cirrhosis; hypertrophic scarring;
 KW atherosclerosis; gene therapy; Duchenne's muscular dystrophy;
 KW insulin-dependent diabetes mellitus; Huntington's; Parkinson's;
 KW Alzheimer's disease; paralysis; cerebellar atrophy; cytosstatic;
 KW neuroprotective; nootropic; antiarthritic; vulnery;
 KW antiatherosclerotic; antidiabetic; hepatotropic; PPM; PR family member.
 XX
 OS Homo sapiens.
 XX
 FN US2003049623-A1.
 XX
 PD 13-MAR-2003.
 XX
 PF 18-JUL-2001; 2001US-00910478.
 XX
 PR 18-JUL-2001; 2001US-00910478.
 XX
 PA (HUAN/) HUANG S.
 XX
 XX Huang S;
 XX
 DR WPI; 2003-567062/53.
 DR N-PSDB; ADF50278.
 XX
 PT New PR/SET-domain containing nucleic acids (which encodes PPM/SET) and
 PT polypeptides, useful for preventing or treating cancers, scleroderma,
 PT arthritis, keloids, atherosclerosis, Huntington's disease or Alzheimer's
 PT disease.
 XX
 PS Claim 21; SEQ ID NO 4; 83pp; English.
 XX
 CC This invention relates to novel PPM (PR family member)/SET (Su(var)3-9,
 CC Enhancer-of-Zeste and Trithorax) nucleotides and encoded polypeptides
 CC thereof that are implicated in proliferative disorders such as cancer.
 CC Specifically, it refers to a functional fragment of the PPM/SET tumour
 CC suppressor gene that encodes a PR, SET, PRAZ or PKZL domain. The present
 CC invention describes the identification and characterization of additional
 CC PR/SET-domain family members that can be used as regulators of cell
 CC proliferation and furthermore to treat, prevent or diagnose
 CC hyperproliferative disorders including scleroderma, arthritis, alcoholic

CC liver cirrhosis, hypertrophic scarring and atherosclerosis. Through gene
CC therapy, these polynucleotides can be used to enhance proliferation of
CC normal cells without rendering the cells cancerous and as such they are
CC particularly useful for treating Duchenne's muscular dystrophy, insulin-
CC dependent diabetes mellitus, Huntington's, Parkinson's, Alzheimer's
CC disease, paralysis, or cerebellar atrophy. Accordingly, these
CC compositions can be described as cytostatic, neuroprotective, or
CC antiarthritic, vulnary, antiatherosclerotic, antidiabetic or
CC hepatotropic. This polypeptide sequence is the human PFM7 protein
CC (encoded by a PFM/SET gene localised to chromosome 11q25) of the
XX invention.
SQ Sequence 1061 AA;

Query Match 11.2%; Score 77.5; DB 7; Length 1061;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 36; Conservative 17; Mismatches 42; Indels 31; Gaps 6;

QY 18 HILQ-----IRKLR-----HRLSNFPRPLPGILAPETVLLPFCYKVRKKEV 59
Db 704 HILKNHGAELPPSIRKLRPAGGEPDMLSTHTQTGTIATPPVCCPHCSQYSSKTKM 763
QY 60 KR--SOKATEFIDYSIEQSHAILTPLOTH-LTMKGSSMKCSLSSEAILFTLTL----- 111
Db 764 VQHIRKKHPEFAQLS-----NTIHTPLTTAVISATPAVLTTDSATGETVVTDDLQAMT 818
QY 112 QLQTQL 117
Db 819 ELSQTL 824

RESULT 12
ID ADO36741 standard; protein; 1061 AA.
XX
AC ADO36741;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human PR-domain containing protein PFM-7.
XX
KW Cytostatic; histone methyltransferase; gene therapy;
KW PFM/SET protein binding motif; cell growth modulator;
KW histone methyltransferase activity; proliferative disorder; cancer;
KW PR-domain; PFM-7.
XX
OS Homo sapiens.
XX
PN US2004014192-A1.
XX
PD 22-JAN-2004.
XX
PF 18-JUL-2002; 2002US-00200012.
XX
PR 18-JUL-2002; 2002US-00200012.
XX
PA (HUAN/) HUANG S.
XX
PI Huang S;
XX
DR WPI; 2004-121568/12.
XX
CC The invention describes an isolated nucleic acid molecule comprising a
CC sequence encoding a PFM/SET protein binding motifs polypeptides
CC comprising a 689, 1061, 367, 717, 571, 951, 1257, 720, or 770 amino acid
CC sequence (PI-P9), given in the specification. Also described are: a
CC vector comprising the isolated nucleic acid molecule; a host cell

CC comprising the vector; an oligonucleotide comprising at least 17
CC contiguous nucleotides of the nucleic acid molecule; a primer pair for
CC detecting PFM/SET nucleic acid molecule, comprising two isolated
CC oligonucleotides; detecting PFM/SET nucleic acid molecule in a sample;
CC modulating cell growth by introducing the vector into a host cell, and
CC expressing the encoded PFM/SET polypeptide in an amount effective to
CC modulate growth of the cell; isolating PFM/SET polypeptide by growing the
CC host cell under conditions appropriate for the expression of the
CC polypeptide; the isolated PFM/SET polypeptide, or its functional fragment
CC; an isolated immunogenic PFM/SET peptide, comprising at least 8
CC contiguous amino acids of PI-P9; an antibody or its antigen-binding
CC fragment that specifically binds to PFM/SET polypeptide; detecting
CC PFM/SET polypeptide in a sample; and screening for a compound that
CC modulates PFM/SET polypeptide histone methyltransferase activity by
CC contacting the polypeptide with one or more candidate compounds, and
CC determining histone methyltransferase activity of the contacted
CC polypeptide. The nucleic acid molecules, polypeptides and methods are
CC useful for diagnosing, prognosing, preventing and treating proliferative
CC disorders, e.g. cancer. This is the amino acid sequence of PR-domain
XX containing polypeptide PFM-7.
SQ Sequence 1061 AA;

Query Match 11.2%; Score 77.5; DB 8; Length 1061;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 36; Conservative 17; Mismatches 42; Indels 31; Gaps 6;

QY 18 HILQ-----IRKLR-----HRLSNFPRPLPGILAPETVLLPFCYKVRKKEV 59
Db 704 HILKNHGAELPPSIRKLRPAGGEPDMLSTHTQTGTIATPPVCCPHCSQYSSKTKM 763
QY 60 KR--SOKATEFIDYSIEQSHAILTPLOTH-LTMKGSSMKCSLSSEAILFTLTL----- 111
Db 764 VQHIRKKHPEFAQLS-----NTIHTPLTTAVISATPAVLTTDSATGETVVTDDLQAMT 818
QY 112 QLQTQL 117
Db 819 ELSQTL 824

RESULT 13
ID ABB58124 standard; protein; 1144 AA.
XX
AC ABB58124;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 1164.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL02227.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.
XX Disclosure; SEQ ID NO 1164; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1144 AA;

Query Match 11.2%; Score 77.5; DB 4; Length 1144;
Best Local Similarity 30.0%; Pred. No. 15;
Matches 30; Conservative 12; Mismatches 33; Indels 25; Gaps 5;

QY 15 NIUIHLOIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKK-----EKVKRSQKATEF 68
Db 718 NVVHELQ-RKLRHVSNNL-----KAIDLPFNKFFFLKTRDHRGQKAK--SQIQKF 765

QY 69 IDYSIEQSH-----HAILPLQTHLTMKSGSMKCSSLS 101
Db 766 LNFIEDDHNGSEAIYTPSPSSDHLKQSLSPKSKSF 805

RESULT 14
ID ADS08184
XX ADS08184 standard; protein; 1026 AA.
AC ADS08184;
XX
XX 04-NOV-2004 (first entry)
XX
DE Staphylococcus epidermis polypeptide seqid 7479.
XX
XX antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW recombinant expression vector; infection; computer readable medium;
KW computer based system.
XX
OS Staphylococcus epidermidis.
XX
XX US2004147734-A1.
XX
XX 29-JUL-2004.
XX
XX 01-DEC-2003; 2003US-00724972.
XX
XX 08-NOV-1997; 97US-0064964P.
XX 13-AUG-1998; 98US-00134001.
XX 29-NOV-1999; 99US-00450969.
XX
XX (DOUC/) DOUCETTE-STAMM L.
XX (BUSH/) BUSH D.
XX
XX Doucette-Stamm L, Bush D;
XX
XX WPI; 2004-580138/56.
XX N-PSDB; ADS04412.
XX
XX New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.
XX
XX Claim 17; SEQ ID NO 7479; 741pp; English.
XX
XX The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any

CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC given in the specification. Also described are: a recombinant expression
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an S. epidermidis polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for S. epidermidis infection; a recombinant or substantially
CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
CC infection. This is the amino acid sequence of a S. epidermis protein of
CC the invention.
XX
SQ Sequence 1026 AA;

Query Match 11.1%; Score 77; DB 8; Length 1026;
Best Local Similarity 20.6%; Pred. No. 15;
Matches 35; Conservative 33; Mismatches 54; Indels 48; Gaps 8;

QY 7 NQKYED-MNIIHILQIR-----KLRHLSNFPRLPGILAPETVL---LPFCYKV 52
Db 782 NQALEDEMHTVTKIVELSRQARKNADLKIKQPLSKM-----VIRPNSQLNLSFLPNYYSI 836

QY 53 FRKEKVR---SQKATEFIDYSIEQSHAI-----LTPLQTHLTMKG 92
Db 837 IKDELNIKNIELTDNIDYITVELKLNFSYVGPGLGNKTKNIQTLDLSLSEYDKKSLIES 896

QY 93 SSMKCSSLSSEA-----ILFTLTQLQTGLLECCLLYLSKTIHPQII 135
Db 897 NNFK--SLSSDAELTKDDFIKTLPKDSYQLSENDVCVILLDKNLSPELI 944

RESULT 15
ABM69043
ID ABM69043 standard; protein; 322 AA.
XX
XX AC ABM69043;
XX
XX 20-NOV-2003 (first entry)
XX
XX Photorhabdus luminescens protein sequence #2140.
XX
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
XX Photorhabdus luminescens.
XX
XX WO200294867-A2.
XX
XX 28-NOV-2002.
XX
XX 07-FEB-2002; 2002WO-IB003040.
XX
XX 07-FEB-2001; 2001FR-00001659.
XX
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;

Search completed: September 9, 2005, 13:00:51
Job time : 62.5 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:25:33 ; Search time 17.5 Seconds
(without alignments)
575.864 Million cell updates/sec

Title: US-09-974-546C-84

Perfect score: 694

Sequence: 1 MRAFLRNQKVEDMNIHIL.....TLGLECCLYLSKTIHPQII 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1 pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	135	3	US-09-097-199-84
2	694	100.0	135	3	US-09-097-199-86
3	694	100.0	157	4	US-09-949-016-9182
4	103	14.8	21	3	US-09-097-199-56
5	72	10.4	636	4	US-09-614-912-92
6	71.5	10.3	634	4	US-09-614-912-88
7	70.5	10.2	438	4	US-09-489-039A-11150
8	70.5	10.2	634	4	US-09-614-912-100
9	69.5	10.0	1294	3	US-09-473-717-2
10	69.5	10.0	1294	4	US-09-949-016-6023
11	69.5	10.0	1353	3	US-09-398-193-99
12	68.5	9.9	2470	4	US-08-265-967C-2
13	68.5	9.9	2470	4	US-08-305-790B-3
14	68	9.8	496	4	US-10-146-704-3
15	68	9.8	639	4	US-09-949-016-6812
16	68	9.8	652	4	US-09-949-016-7323
17	68	9.8	993	4	US-09-949-016-10335
18	67.5	9.7	156	4	US-09-543-681A-7434
19	67	9.7	399	4	US-09-543-681A-4933
20	67	9.7	517	3	US-09-457-040B-14
21	67	9.7	773	2	US-08-484-101B-42
22	67	9.7	773	2	US-08-484-101B-44
23	67	9.7	773	3	US-08-714-524D-42
24	67	9.7	773	3	US-08-714-524D-44
25	66.5	9.6	580	4	US-09-107-532A-5734
26	66.5	9.6	1784	3	US-09-040-738-2
27	66.5	9.6	1784	3	US-08-652-426A-2

28	66.5	9.6	1813	4	US-09-949-016-8283	Sequence 8283, Ap
29	66.5	9.6	1813	4	US-09-949-016-8284	Sequence 8284, Ap
30	66.5	9.6	1813	4	US-09-949-016-8285	Sequence 8285, Ap
31	66	9.5	267	4	US-09-270-767-46386	Sequence 46386, A
32	66	9.5	348	4	US-09-198-452A-922	Sequence 922, App
33	66	9.5	348	4	US-09-438-185A-856	Sequence 856, App
34	66	9.5	611	4	US-09-949-016-7139	Sequence 7139, Ap
35	65.5	9.4	598	4	US-09-538-092-151	Sequence 151, App
36	65.5	9.4	3218	1	US-08-764-100-27	Sequence 27, Appl
37	65	9.4	75	4	US-09-621-976-6155	Sequence 6155, Ap
38	65	9.4	258	4	US-09-270-767-33067	Sequence 33067, A
39	65	9.4	258	4	US-09-270-767-48284	Sequence 48284, A
40	65	9.4	288	1	US-08-375-709-9	Sequence 9, Appli
41	65	9.4	288	1	US-08-752-929-9	Sequence 9, Appli
42	65	9.4	288	3	US-09-090-793-6	Sequence 6, Appli
43	65	9.4	288	4	US-09-231-899-6	Sequence 6, Appli
44	64.5	9.3	236	4	US-09-107-532A-6873	Sequence 6873, Ap
45	64.5	9.3	315	3	US-09-134-001C-5446	Sequence 5446, Ap

ALIGNMENTS

RESULT 1

US-09-097-199-84
; Sequence 84, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-097-199-84

Query Match 100.0%; Score 694; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 4.7e-77;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MF	AFLRNQKYEDMNI	I	H	I	Q	I	K	L	R	H	R	S	N	F	P	G	I	L	A	P	E	T	V	L	L	P	F	C	V	K	V	P	R	K	E	K	V	K	60
Db	1	MF	AFLRNQKYEDMNI	I	H	I	Q	I	K	L	R	H	R	S	N	F	P	G	I	L	A	P	E	T	V	L	L	P	F	C	V	K	V	P	R	K	E	K	V	K	60
Qy	61	RSQK	ATEFDYIS	E	O	S	H	A	I	L	P	T	Q	T	H	L	T	M	K	G	S	M	K	C	S	S	S	S	E	A	I	L	P	T	L	T	L	T	L	120	
Db	61	RSQK	ATEFDYIS	E	O	S	H	A	I	P	T	Q	T	H	L	T	M	K	G	S	M	K	C	S	S	S	S	E	A	I	L	P	T	L	T	L	T	L	120		

Qy 121 CCLLYLSKTIHPQII 135
 | | | | | | | | | |
Db 121 CCLLYLSKTIHPQII 135

```

RESULT 2
US-09-097-199-86
; Sequence 86, Application US/09037199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210

```

```
Query Match      100.0%; Score 694; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 4.7e-77;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	MAEFLRNQKYEDMNNIIHILQIRKLRLHLSNFPRLPGILAPETVLLPFCYKVPKRKEKVK	60
Db	1	MAEFLRNQKYEDMNNIIHILQIRKLRLHLSNFPRLPGILAPETVLLPFCYKVPKRKEKVK	60
Qy	61	RSQKATEFDYDIEQSHAILTPQLTHLMKGSSMKCSSLSSEAILFTLLQLTQTGLLE	120
Db	61	RSQKATEFDYDIEQSHAILTPQLTHLMKGSSMKCSSLSSEAILFTLLQLTQTGLLE	120
Qy	121	CCLLYLSKTIHPQII	135

Db 121 CCLLYLSKTIHPQII 135

```

RESULT 3
US-09-949-016-9182
; Sequence 9182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9182
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9182

```

```

Query Match      100.0%; Score 694; DB 4; Length 157;
Best Local Similarity 100.0%; Pred. No. 5.8e-77;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

[illegible]

```

RESULT 4
US-09-097-199-56
; Sequence 56, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:59:09 ; Search time 52.5 Seconds
(without alignments)
1014.252 Million cell updates/sec

Title: US-09-974-546C-84

Perfect score: 694

Sequence: 1 MRAFLNQKVEDMNIHIL.....TLGLECLLYLSTKTHPQII 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1777461 seqs, 394431504 residues

Total number of hits satisfying chosen parameters: 1777461

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/PCTUS_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	135	10	US-09-974-546-84
2	694	100.0	135	10	US-09-974-546-86
3	103	14.8	21	10	US-09-974-546-56
4	90	13.0	371	15	US-10-369-493-20398
5	90	13.0	381	15	US-10-369-493-1334
6	81	11.7	1478	16	US-10-425-115-362997
7	79	11.4	264	15	US-10-424-599-174106
8	79	11.4	1368	16	US-10-425-115-324108
9	79	11.4	1444	16	US-10-425-115-324095
10	78	11.2	289	15	US-10-424-599-148597
11	78	11.2	3229	16	US-10-425-115-324093

12	77.5	11.2	1061	15	US-10-200-012-4	Sequence 4, Appli
13	77	11.1	1026	18	US-10-724-972A-7479	Sequence 7479, Ap
14	77	11.1	1766	16	US-10-425-115-324098	Sequence 324098,
15	75	10.8	633	15	US-10-282-122A-70203	Sequence 70203, A
16	75	10.8	1259	16	US-10-425-115-324122	Sequence 324122,
17	75	10.8	3749	16	US-10-425-115-324125	Sequence 324125,
18	74.5	10.7	388	15	US-10-424-599-169767	Sequence 169767,
19	74.5	10.7	404	18	US-10-954-778-97	Sequence 97, Appli
20	74	10.7	1353	16	US-10-425-115-362744	Sequence 262744,
21	74	10.7	1605	16	US-10-425-115-362998	Sequence 262998,
22	74	10.7	1676	16	US-10-425-115-238569	Sequence 238569,
23	74	10.7	1708	16	US-10-425-115-324103	Sequence 324103,
24	74	10.7	1752	16	US-10-425-115-238562	Sequence 238562,
25	74	10.7	1754	16	US-10-425-115-324104	Sequence 324104,
26	73	10.5	1143	16	US-10-425-115-238564	Sequence 238564,
27	73	10.5	1517	16	US-10-425-115-324099	Sequence 324099,
28	72.5	10.4	3365	16	US-10-739-930-6537	Sequence 6537, Ap
29	72	10.4	329	15	US-10-424-599-150082	Sequence 150082,
30	71.5	10.3	407	15	US-10-425-114-69882	Sequence 69882, A
31	71.5	10.3	426	16	US-10-425-115-327356	Sequence 327356,
32	71.5	10.3	676	15	US-10-424-599-267862	Sequence 267862,
33	71	10.2	334	15	US-10-425-114-63742	Sequence 63742, A
34	71	10.2	1214	16	US-10-425-115-243074	Sequence 243074,
35	71	10.2	1595	16	US-10-425-115-238559	Sequence 238559,
36	71	10.2	1715	16	US-10-425-115-324123	Sequence 324123,
37	71	10.2	1924	9	US-09-866-557A-2	Sequence 2, Appli
38	71	10.2	1924	11	US-09-858-862-2	Sequence 2, Appli
39	71	10.2	1924	14	US-10-055-797-2	Sequence 2, Appli
40	71	10.2	1924	15	US-10-350-798-2	Sequence 2, Appli
41	70.5	10.2	269	15	US-10-424-599-264790	Sequence 264790,
42	70.5	10.2	634	17	US-10-876-086-27	Sequence 27, Appli
43	70	10.1	206	16	US-10-408-765A-1698	Sequence 1698, Ap
44	70	10.1	272	16	US-10-437-963-125982	Sequence 125982,
45	70	10.1	327	15	US-10-425-114-63743	Sequence 63743, A

ALIGNMENTS

RESULT 1
US-09-974-546-84

; Sequence 84, Application US/09974546
; Publication No. US20030050470A1

; GENERAL INFORMATION:

; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert

; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,

; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE

; NUMBER OF SEQUENCES: 87

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P. O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: USA

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/974,546

; FILING DATE: 10-Oct-2001

; CLASSIFICATION: Unknown

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/097,199

; FILING DATE: 1998-06-12

; ATTORNEY/AGENT INFORMATION:

; NAME: Nakashima, Richard A.

; REGISTRATION NUMBER: P-42,023

```
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 84:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 135 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-974-546-84

Query Match 100.0%; Score 694; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 9.1e-72;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLNQYEDMHNIIHLQIRKLHRLSNPPRLPGILAPETVLLPFCYKVRKKEVK 60
Db 1 MRAFLNQYEDMHNIIHLQIRKLHRLSNPPRLPGILAPETVLLPFCYKVRKKEVK 60

QY 61 RSQKATEFDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTGLG 120
Db 61 RSQKATEFDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTGLG 120

QY 121 CCLLYLSKTIHPQII 135
Db 121 CCLLYLSKTIHPQII 135

RESULT 2
US-09-974-546-86
; Sequence 86, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Velttri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/974,546
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/097,199
; FILING DATE: 1998-06-12
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 474-7577
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; QUERY MATCH 14.8%; Score 103; DB 10; Length 21;
; Best Local Similarity 100.0%; Pred. No. 0.00014;
; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 RKKEKRSQKATEFDYSIE 74
```

```
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

Query Match 100.0%; Score 694; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 9.1e-72;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLNQYEDMHNIIHLQIRKLHRLSNPPRLPGILAPETVLLPFCYKVRKKEVK 60
Db 1 MRAFLNQYEDMHNIIHLQIRKLHRLSNPPRLPGILAPETVLLPFCYKVRKKEVK 60

QY 61 RSQKATEFDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTGLG 120
Db 61 RSQKATEFDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTGLG 120

QY 121 CCLLYLSKTIHPQII 135
Db 121 CCLLYLSKTIHPQII 135

RESULT 3
US-09-974-546-56
; Sequence 56, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Velttri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/974,546
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/097,199
; FILING DATE: 1998-06-12
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-974-546-56

Query Match 14.8%; Score 103; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 RKKEKRSQKATEFDYSIE 74
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 13:01:39 ; Search time 14.5 Seconds
(without alignments)
895.811 Million cell updates/sec

Title: US-09-974-546C-84

Perfect score: 694

Sequence: 1 MRAFLRNQKYEDMHNIIHL.....TLGLECLLYLSKTIHPQII 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	13.0	381	2 F71196	probable hexosyltr
2	82.5	11.9	645	2 F86174	hypothetical prote
3	80	11.5	495	1 MNXRSA	nonstructural prot
4	78.5	11.3	3147	2 T21328	hypothetical prote
5	78	11.2	523	2 D64555	conserved hypotet
6	74.5	10.7	172	2 A58456	X protein - human
7	74.5	10.7	263	2 A82069	hypothetical prote
8	72.5	10.4	796	2 T43782	hypothetical prote
9	71.5	10.3	227	2 B84040	ABC transporter (A
10	71.5	10.3	451	2 D70045	two-component sens
11	70.5	10.2	1064	2 T40751	isoleucyl-trna syn
12	70	10.1	404	2 T06761	hypothetical prote
13	70	10.1	982	1 A44831	phosphoenolpyruvat
14	70	10.1	982	2 AE2413	phosphoenolpyruvat
15	69.5	10.0	1398	2 H71606	hypothetical prote
16	69	9.9	154	2 T27983	hypothetical prote
17	69	9.9	743	2 T00634	hypothetical prote
18	69	9.9	1850	2 AC1917	serine/threonine k
19	68.5	9.9	503	1 S11338	steroid libeta-mon
20	68.5	9.9	2470	2 S57085	1-phosphatidylinos
21	68	9.8	379	2 C75006	hypothetical prote
22	68	9.8	414	1 H64203	histidine-tRNA lig
23	68	9.8	456	2 A31986	glucose transporte
24	68	9.8	639	2 T13151	adaptor protein CM
25	68	9.8	5126	2 S40450	ryanodine receptor
26	67.5	9.7	619	2 B87682	hypothetical prote
27	67.5	9.7	833	2 T82385	ORF MSV224 probabl
28	67.5	9.7	1278	2 A71609	probable secreted
29	67.5	9.7	3225	2 D81702	adherence factor T

30	67	9.7	126	2 T03757	hypothetical prote
31	67	9.7	213	2 E95506	hypothetical prote
32	67	9.7	317	2 B82440	hypothetical prote
33	67	9.7	347	2 T01452	translation initia
34	67	9.7	517	2 A54099	protein kinase Dar
35	67	9.7	587	2 S37927	hypothetical prote
36	66.5	9.6	244	2 AC3341	precorrin-2 C20-me
37	66.5	9.6	447	2 T21716	hypothetical prote
38	66.5	9.6	635	2 T08537	ethylene receptor
39	66.5	9.6	859	2 H70327	DNA mismatch repal
40	66.5	9.6	908	2 T22376	hypothetical prote
41	66.5	9.6	1784	2 A49420	tuberos sclerosi
42	66	9.5	344	2 D86597	outer membrane pro
43	66	9.5	344	2 H72027	major outer membra
44	66	9.5	446	2 A42029	transcription fact
45	66	9.5	472	2 F97147	probable Fe-S oxid

ALIGNMENTS

RESULT 1

F71196
probable hexosyltransferase (EC 2.4.1.1-) PH1844 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: F71196
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic ar
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: F71196
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-381 <KAW>
A:Cross-references: UNIPROT:O59512; GB:AP000007; NID:G3236134; PID:BAA30965.1; PID:G325E
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1844
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match	13.0%;	Score 90;	DB 2;	Length 381;
Best Local Similarity	26.8%;	Pred. No. 0.16;		
Matches	34;	Conservative	23;	Mismatches 38; Indels 32; Gaps 7;
Qy	13	MNIIHILQIRKLRHL	----	SNFP-----RLPGILAPETVLLPFCYKVF 53
Db	24	MNLL--AIKLRGHEVGIVNRTGKEELKRYGIELIKIPGIISP	----	FLDV 73
Qy	54	RKKEVKYRSQKATEFI-DYSIEQSHALTPLOTHLTKWGSSM-KCSSLSSEAILFTLTL 111		
Db	74	NLTGYLKSSEELNEFKDFDIHSHH-FTPLSLKALKAGKNKGTLLTTHSISPAHES 132		
Qy	112	QLTQTILG 118		
Db	133	KLMDTLG 139		

RESULT 2

F86174
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: F86174
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ket, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F86174

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-645 <STO>

A;Cross-references: UNIPROT:P93825; GB:AE005172; NID:g2341041; PIDN:AAB70445.1; GSPDB:GN

C;Genetics:

A;Map position: 1

Query Match 11.9%; Score 82.5; DB 2; Length 645;

Best Local Similarity 23.5%; Pred. No. 1.7;

Matches 31; Conservative 26; Mismatches 46; Indels 29; Gaps 5;

QY 25 LHRHSNF---PRLPGILAPETV-----LLPFCYKVRKKEKVKRQ 63

DB 99 MTHLLAGFTYGPMPWMTAVTVFKMLTGIVSFLTALSLSVTLPLLLKA--KVREPLMSK 156

QY 64 KATEFDYIEQSHHAILTPLOTHLTMKGSSMKC--SSLSSAILFTLTQLTQTLGLECC 122

DB 157 KYRE-----LDREVGLIMQKQTSLSHVRMLTKIRTSLSRHTLYTLVELSKTLGLKNC 211

QY 123 LLYLSKTIHPQI 134

DB 212 AVWIPNEIKTEM 223

RESULT 3

MNXRSA

nonstructural protein NCVP2 - simian rotavirus SALL

N;Alternate names: nonstructural protein NS53

C;Species: simian rotavirus Sall

C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004

C;Accession: S08215

R;Mitchell, D.B.; Both, G.W.

Virology 174, 618-621, 1990

A;Title: Conservation of a potential metal binding motif despite extensive sequence divergence

A;Reference number: S08215; MUID:90163231; PMID:2154894

A;Accession: S08215

A;Molecule type: Genomic RNA

A;Residues: 1-495 <MIT>

A;Cross-references: UNIPROT:PI5687; EMBL:X14914; NID:g61889; PIDN:CAA33039.1; PTD:g61890

C;Genetics:

A;Map position: segment 5

C;Superfamily: bovine rotavirus nonstructural protein NCVP2

C;Keywords: nonstructural protein; zinc finger

Query Match 11.5%; Score 80; DB 1; Length 495;

Best Local Similarity 25.3%; Pred. No. 2.3;

Matches 24; Conservative 22; Mismatches 33; Indels 16; Gaps 4;

QY 2 RAFLRNQKVEDMNIHILQIRKLRLNSFPRLPGILAPETVLLPFCYKVRKKEKVKR 61

DB 71 RCFLDNEP-----HLKRLTKVHPITK-DKLCQIIDLYNIIFINDKVRKPERMIK 121

QY 62 SOKATEFDYIEQSHHAILTPLOTHLTMKGSSMK 96

DB 122 QRKCRN--QTKIEWNHLLP-----ITLNAAFK 149

RESULT 4

T21328

hypothetical protein F25C8.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T21328

R;Matthews, L.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19407

A;Accession: T21328

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-3147 <WIL>

A;Cross-references: UNIPROT:O9XV66; EMBL:Z81512; PIDN:CAB04172.1; GSPDB:GN00023; CBSP:F2E

A;Experimental source: clone F25C8

C;Genetics:

A;Gene: CESP:F25C8.3

A;Map position: 5

A;Introns: 32/2; 57/2; 73/3; 113/3; 157/3; 366/1; 456/1; 509/1; 638/1; 773/2; 848/1; 902/

/3; 2510/2; 2737/3; 2827/1; 2902/3; 3093/3; 3121/1

C;Superfamily: Caenorhabditis elegans hypothetical protein F25C8.3

Query Match 11.3%; Score 78.5; DB 2; Length 3147;

Best Local Similarity 28.1%; Pred. No. 26;

Matches 41; Conservative 17; Mismatches 67; Indels 21; Gaps 6;

QY 2 RAFLRN-----QKVEDMNIHILQIRKLRLNSFPRLPGILAPETVLLPFCYKVKR 54

DB 1571 RADKKNLADNNQAKQQAQALRKSIHARQSTAVPRRESAMVGQPE-FASKAIRKMLMEKMQQ 1629

QY 55 KKEKVKRQKATE--FIDYSIEQSHHAILTPLOTHLTMKGSSMKCSSLSSSEAILFTLTQ 112

DB 1630 EKEKEKEKEKDALKKQSVQDHSSTDTBEDAQLPEKNKPM-----TYLSLSVLQ 1682

QY 113 LTQ---TGLGECCLLYLSKTIHPQII 135

DB 1683 LVHSPISSVLKCCLL-LSVEQHKQMI 1707

RESULT 5

D64555

conserved hypothetical integral membrane protein HP0284 - Helicobacter pylori (strain 26t

C;Species: Helicobacter pylori

C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004

C;Accession: D64555

R;Tombs, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.N

A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A;Reference number: A64520; MUID:97394467; PMID:9252185

A;Accession: D64555

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-523 <TOM>

A;Cross-references: UNIPROT:O25059; GB:AE000547; GB:AE000511; NID:g2313377; PIDN:AAD07351

Query Match 11.2%; Score 78; DB 2; Length 523;

Best Local Similarity 23.5%; Pred. No. 3.9;

Matches 24; Conservative 16; Mismatches 30; Indels 32; Gaps 3;

QY 5 LRNQKVEDMNIHILQIRKLRLNS-----NPPRLPGILA 40

DB 71 LKNKASKD-----LKISTLEHTLKALSSQOKMPESGVNPFKDLIERNPINPIAN 124

QY 41 PETVL--LPFCYKVRKKEKVKRSQKATEFIDYSIEQSHHAI 80

DB 125 PLAIIDGISFIKSMHLKHENLKNQTALEEVLRLLDQKHQL 166

RESULT 6

A58456

X protein - human hepatitis B virus

C;Species: human hepatitis B virus

C;Date: 30-Sep-1993 #sequence_revision 21-Jan-1997 #text_change 09-Jul-2004

C;Accession: A58456

R;Zeng, M.; Huang, B.R.; Cai, L.W.; Pan, G.Z.

Chinese Biochem. J. 12, 22-26, 1996

A;Title: The sequence analysis of the hepatitis B virus (HBV) X gene and its expression i

A;Reference number: JCI1331

A;Accession: A58456

A;Molecule type: DNA

A;Residues: 1-172 <ZEN>

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	694	100.0	135	2	Q9GZY1	homo sapien
2	90	13.0	381	2	O59512	pyrococcus
3	82.5	11.9	645	2	P93825	arabidopsis
4	81	11.7	451	2	Q7SZX8	brachydanio
5	80	11.5	414	2	O99FX6	simian rota
6	80	11.5	479	2	Q99FX4	simian rota
7	80	11.5	495	1	VN53 ROTs1	simian 11 r
8	80	11.5	496	2	O99FX5	simian rota
9	80	11.5	496	2	Q99FX7	simian rota
10	78.5	11.3	3175	2	Q7JKT8	caenorhabdi
11	78	11.2	523	2	O25059	helicobacte
12	78	11.2	1646	2	O86820	potato aucu
13	77.5	11.2	358	2	Q8TZU8	pyrococcus
14	77.5	11.2	931	2	Q9NX14	homo sapien
15	77.5	11.2	1023	2	O863Z2	homo sapien
16	77.5	11.2	1661	1	PRDA_HUMAN	homo sapien
17	77.5	11.2	1061	2	Q9ULI9	homo sapien
18	77.5	11.2	1117	2	Q874E8	homo sapien
19	77.5	11.2	1117	2	Q9W3N0	drosophila
20	77	11.1	495	1	VN53 ROTSP	simian 11 r
21	77	11.1	1024	1	SVIP_STAAU	staphylococ
22	76.5	11.0	321	2	Q7N442	photornhabu
23	76.5	11.0	398	2	Q6LRR8	photobacter
24	76.5	11.0	2259	1	YCZ2_PHYPA	physcomitre
25	75	10.8	629	2	O6GJ08	staphylococ
26	74.5	10.7	160	2	Q6S1F8	uncultured
27	74.5	10.7	163	2	O6SGJ3	uncultured
28	74.5	10.7	172	2	Q7MW22	hepatitis b
29	74.5	10.7	263	2	Q9KP79	vibrio chol
30	74.5	10.7	404	2	Q7PNX8	anopheles g
31	74	10.7	3184	2	Q9XV66	caenorhabdi

```

RP SEQUENCE FROM N.A.
RA Theologis;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Hua J., Sakail H., Nourizadeh S., Chen Q.G., Bleecker A.B.,
RL Ecker J.R., Meyerowitz E.M.;
RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Sakai H., Hua J., Chen Q.G., Chang C., Medrano L.J., Bleecker A.B.,
RL Meyerowitz E.M.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC000104; AAB70445.1; -.
DR EMBL; AF047976; AAC62209.1; -.
DR PIR; F86174; F86174.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR003661; His_kinA_N.
DR InterPro; IPR011052; Prot_aml_inh.
DR Pfam; PF01590; GAF; 1.
DR Pfam; PF00512; HsKa; 1.
DR SMART; SM00065; GAF; 1.
DR SMART; SM00386; HsKa; 1.
DR Receptor.
KW
SQ SEQUENCE 645 AA; 72192 MW; D37ABE3A0939D6ED CRC64;
Query Match 11.9%; Score 82.5; DB 2; Length 645;
Best Local Similarity 23.5%; Pred. No. 17;
Matches 31; Conservative 26; Mismatches 46; Indels 29; Gaps 5;
QY 25 LRHRLSNF---PRLPGILAPETV-----LLPCYKVKRKEKVKRSQ 63
DB 99 MTHLLAGTYGPHWPWMTAVTFKMLTGIVSFLTALSIVTLPLLLKA--KVREPMLSK 156
QY 64 KATEFDYSIEQSHAILTPLOTHITMKGSSMKC--SSLSEAILFTLTQLTQTGLECC 122
DB 157 KTRF-----LDREVGIIKQTETSLHVRMLTKIRTSIDRRHTILYTLVELSKTILGLKNC 211
QY 123 LHLXSKTIHPQI 134
DB 212 AVWIPNEIKTEM 223
RESULT 4
ID Q7SZX8 PRELIMINARY; PRT; 451 AA.
AC Q7SZX8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE SI:zK13A21.9 (Novel protein similar to vertebrate microphthalmia-
DE associated transcription factor (MITF) and zebrafish transcription
DE factor binding to IGHM enhancer 3a (Tfe3a)).
DE Name=opnllw2; Synonyms=SI:zK13A21.9;
DE OS Brachydanio rerio (Zebrafish) (Danio rerio).
DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
DE Cyprinidae; Danio.
DE NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pandian R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
DR EMBL; AL844847; CAE30419.1; -.
DR HSP; P22415; 1AN4.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 9, 2005, 06:16:18 ; Search time 119.463 Seconds
(without alignments)
16226.276 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gacctaataatcgaggt.....agtaaaactcccccaacc 2506

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-O=/cpn2.1/uspto.spool/US09974546/runat.07092005.174357.19482/app_query.fasta_1.4942
-DB=A_Geneseq_16Dec04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=pct -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODES=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CGN 1.1 232 @runat.07092005.174357.19482 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	694	15.9	135	3 AAY59296	Aay59296 Prostate
2	694	15.9	135	3 AAY59295	Aay59295 Prostate
3	694	15.9	135	4 AAU02174	Aau02174 Biomarker
4	694	15.9	135	4 AAU02175	Aau02175 Biomarker
5	313.5	7.2	129	8 ADRO8959	Adr08959 Human pro
6	308.5	7.1	110	3 AAB43292	Aab43292 Human ORF
7	304.5	7.0	95	4 AAO06985	Aao06985 Human pol
8	302	7.0	118	8 ADQ67624	Adq67624 Novel hum
9	300.5	6.9	119	5 ADK37016	Adk37016 Novel hum
10	298.5	6.9	100	4 AAB92614	Aab92614 Human pro

11	296.5	6.8	7036	6 AAE35264	Aae35264 Human P45
12	295.5	6.8	126	5 ABG65578	Abg65578 Human bre
13	293.5	6.8	98	5 ADG79629	Adg79629 Human sec
14	293.5	6.8	341	4 AAU27917	Aau27917 Human con
15	291.5	6.7	102	4 AAO08307	Aao08307 Human pol
16	291	6.7	119	4 AAM83429	Aam83429 Human imm
17	290.5	6.7	97	8 ADM87815	Adm87815 Human EST
18	290.5	6.7	202	4 AAE11962	Aae11962 Human c-m
19	289	6.6	151	5 AAU76310	Aau76310 Human fas
20	288.5	6.7	129	8 ADQ65903	Adq65903 Novel hum
21	288	6.7	164	4 AAE07113	Aae07113 Human gen
22	287.5	6.6	116	5 ADG79555	Adg79555 Human sec
23	287.5	6.6	116	5 ADG79456	Adg79456 Human sec
24	286.5	6.6	164	8 ADQ66320	Adq66320 Novel hum
25	285	6.6	107	4 AAB94965	Aab94965 Human pro
26	284.5	6.6	144	5 ADK37055	Adk37055 Novel hum
27	284.5	6.6	170	7 ADC86487	Adc86487 Human GPC
28	283.5	6.6	116	6 AAQ26240	Aao26240 MDDT rela
29	283.5	6.6	119	4 AAO06149	Aao06149 Human pol
30	283.5	6.6	128	4 ABB11824	Abb11824 Human sec
31	283.5	6.6	135	7 ADE08953	Ade08953 Novel pro
32	283.5	6.6	135	8 ADS11754	Ads11754 Human the
33	283.5	6.6	152	7 ADB37629	Adb37629 Neural th
34	283	6.5	127	7 ADB37543	Adb37543 Neural th
35	282.5	6.5	99	4 AAG78806	Aag78806 Human neu
36	282.5	6.5	99	5 ADK36872	Adk36872 Novel hum
37	282.5	6.5	119	4 AAU86631	Aau86631 Novel hum
38	282.5	6.5	119	7 ADB59965	Adb59965 Connectiv
39	282.5	6.5	127	7 ADM03873	Adm03873 Human pro
40	282	6.5	129	4 AAB95167	Aab95167 Human pro
41	282	6.5	129	7 ADB37541	Adb37541 Neural th
42	282	6.5	136	7 ADB65293	Adb65293 Human pro
43	281.5	6.5	115	4 AAU32097	Aau32097 Novel hum
44	281.5	6.5	125	8 ADR08586	Adr08586 Human pro
45	281.5	6.5	260	8 ADR09618	Adr09618 Human pro

ALIGNMENTS

RESULT 1
ID AAY59296 standard; peptide; 135 AA.
XX AAY59296;
AC AAY59296;
XX
XX 19-APR-2000 (first entry)
XX
XX Prostate disease marker UC Band #28 amino acid sequence.

Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.

Homo sapiens.

WO9964631-A1.

16-DEC-1999.

11-JUN-1999; 99WO-US013151.

12-JUN-1998; 98US-00097199.

(UROC-) UROCOR INC.

An G, O'hara SM, Ralph D, Veltri RW;
WPI; 2000-116557/10.

N-PSDB; AA287584.

Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
breast and bladder cancer.

Example 5; Page 184-186; 191pp; English.

XX The invention provides nucleic acid markers of prostate, breast and
 CC bladder cancer. The markers are indicators of malignant transformation of
 CC prostate, breast and bladder tissues and are diagnostic of the potential
 CC for metastatic spread of malignant prostate tumours. The nucleic acid can
 CC also be used as targets for therapeutic intervention in prostate cancer,
 CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
 CC markers may be used to design specific probes and primers, for the rapid
 CC analysis of prostate, bladder or breast biopsy samples. The probes and
 CC primers may also be used for in situ hybridization or in situ PCR
 CC detection and diagnosis. They may also be used to identify and isolate
 CC full length gene sequences from various DNA libraries. Antibodies against
 CC the polypeptide products of the markers can be used to treat prostate
 CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
 CC to detect antibodies. The proteins and antibodies can be used in
 CC immunodetection methods for detecting or quantifying the cancers, and for
 CC clinical diagnosis of these cancers. The antibodies may also be used for
 CC radioimaging to quantify and localize the encoded proteins
 XX Sequence 135 AA;

Alignment Scores:
 Pred. No.: 3.14e-73 Length: 135
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 15.86% Indels: 0
 DB: 3 Gaps: 0

US-09-974-546C-85 (1-2506) x AAY59296 (1-135)

QY 99 ATGAGGCGCTCTTAAAGGAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
 Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20
 QY 159 CAGATCAGAAAATTTGAGGCACAGATTAAGTAATCTTCCCAAGGCTACAGGCATTCCTAGCT 218
 Db 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 QY 219 CCAGAACTGTCTCTTACCATTCTGCTACAGGATTTTCGAAAAAGAAAAGTAAAA 278
 Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGlnLysValLys 60
 QY 279 AGAAGTCAAAAGGCAACAGATTCATTGATTTATTCATAGACAGTCACACCATGCAATT 338
 Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 QY 339 CTCACACCTTCGACACACATTTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 398
 Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 QY 399 TCTTCAGAGCCATATTATTCACATTGACTTTGCAGTTAACTCAACCTAGGCTCGGAA 458
 Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 QY 459 TCTGTCTTCTCTACTATTATCCAAAACATATACATCCACAGATCATTA 503
 Db 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

RESULT 2

AA59295
 ID AAY59295 standard; peptide; 135 AA.
 XX
 AC AAY59295;
 XX
 DT 19-APR-2000 (first entry)
 XX
 DE Prostate disease marker UC Band #28 amino acid sequence.
 XX
 KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
 KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
 XX
 OS Homo sapiens.

XX WO9964631-A1.
 PN 16-DEC-1999.
 PD 11-JUN-1999; 99WO-US013151.
 PF 12-JUN-1998; 98US-00097199.
 PR (UROC-) UROCOR INC.
 XX An G, O'hara SM, Ralph D, Veltri RW;
 PI WPI; 2000-116557/10.
 DR N-PSDB; AAZ87583.
 XX Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
 PT breast and bladder cancer.
 XX Example 5; Page 182-183; 191pp; English.
 PS The invention provides nucleic acid markers of prostate, breast and
 CC bladder cancer. The markers are indicators of malignant transformation of
 CC prostate, breast and bladder tissues and are diagnostic of the potential
 CC for metastatic spread of malignant prostate tumours. The nucleic acid can
 CC also be used as targets for therapeutic intervention in prostate cancer,
 CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
 CC markers may be used to design specific probes and primers, for the rapid
 CC analysis of prostate, bladder or breast biopsy samples. The probes and
 CC primers may also be used for in situ hybridization or in situ PCR
 CC detection and diagnosis. They may also be used to identify and isolate
 CC full length gene sequences from various DNA libraries. Antibodies against
 CC the polypeptide products of the markers can be used to treat prostate
 CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
 CC to detect antibodies. The proteins and antibodies can be used in
 CC immunodetection methods for detecting or quantifying the cancers, and for
 CC clinical diagnosis of these cancers. The antibodies may also be used for
 CC radioimaging to quantify and localize the encoded proteins
 XX Sequence 135 AA;

Alignment Scores:
 Pred. No.: 3.14e-73 Length: 135
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 15.86% Indels: 0
 DB: 3 Gaps: 0

US-09-974-546C-85 (1-2506) x AAY59295 (1-135)

QY 99 ATGAGGCGCTCTTAAAGGAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
 Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeu 20
 QY 159 CAGATCAGAAAATTTGAGGCACAGATTAAGTAATCTTCCCAAGGCTACAGGCATTCCTAGCT 218
 Db 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 QY 219 CCAGAACTGTCTCTTACCATTCTGCTACAGGATTTTCGAAAAAGAAAAGTAAAA 278
 Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGlnLysValLys 60
 QY 279 AGAAGTCAAAAGGCAACAGATTCATTGATTTATTCATAGACAGTCACACCATGCAATT 338
 Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 QY 339 CTCACACCTTCGACACACATTTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 398
 Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 QY 399 TCTTCAGAGCCATATTATTCACATTGACTTTGCAGTTAACTCAACCTAGGCTCGGAA 458
 Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 QY 459 TCTGTCTTCTCTACTATTATCCAAAACATATACATCCACAGATCATTA 503
 Db 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 9, 2005, 11:23:58 ; Search time 25.911 Seconds
(without alignments)
14439.486 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gacctaataatagaggt.....agtaaaactccccccacc 2506

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/070974546/runat_07092005_174359_19522/app_query.fasta_1.4942
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0 -L=LOOPCL=0
-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=blosu62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CIGN 1 1 42 @runat_07092005_174359_19522 -NCPV=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBPLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PTCUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	771	17.6	157	4	US-09-949-016-9182
2	694	15.9	135	3	US-09-097-199-84
3	694	15.9	135	3	US-09-097-199-86
4	278.5	6.4	310	4	US-09-800-729-190
5	277.5	6.4	239	4	US-09-800-729-193
6	263	6.1	144	4	US-09-513-999C-6953
7	252.5	5.8	108	4	US-09-513-999C-4433
8	249.5	5.8	375	2	US-08-454-557C-121
9	249.5	5.8	375	2	US-08-340-426D-121
10	249.5	5.8	375	2	US-08-450-673C-121
11	249.5	5.8	375	4	US-09-872-968-2
12	248.5	5.7	96	4	US-09-513-999C-6065

13	240.5	5.5	95	4	US-09-621-976-3964	Sequence 3964, Ap
14	240	5.5	169	4	US-09-663-600A-208	Sequence 208, App
15	235.5	5.4	119	4	US-09-513-999C-7867	Sequence 7867, Ap
16	231	5.3	92	4	US-09-248-796A-16335	Sequence 16335, A
17	230	5.3	99	4	US-09-288-143-168	Sequence 168, App
18	229.5	5.3	112	4	US-09-513-999C-7870	Sequence 7870, Ap
19	229.5	5.3	118	4	US-09-663-600A-114	Sequence 114, App
20	224	5.2	1079	3	US-09-058-483-22	Sequence 22, Appl
21	223	5.2	397	5	PCT-US95-1711A-121	Sequence 121, App
22	221.5	5.1	132	4	US-09-636-215-573	Sequence 573, App
23	221.5	5.1	132	4	US-09-685-166A-573	Sequence 573, App
24	221.5	5.1	132	4	US-09-679-426-573	Sequence 573, App
25	221.5	5.1	132	4	US-09-759-143-573	Sequence 573, App
26	221.5	5.1	132	4	US-09-651-236-573	Sequence 573, App
27	221.5	5.1	135	4	US-09-685-166A-884	Sequence 884, App
28	221.5	5.1	135	4	US-09-679-426-884	Sequence 884, App
29	221.5	5.1	135	4	US-09-759-143-884	Sequence 884, App
30	220	5.1	91	4	US-09-621-976-5929	Sequence 5929, Ap
31	220	5.0	107	4	US-09-513-999C-7879	Sequence 7879, Ap
32	218	5.0	121	4	US-09-513-999C-7874	Sequence 7874, Ap
33	217.5	5.0	84	4	US-09-621-976-4377	Sequence 4377, Ap
34	216	5.0	92	4	US-09-621-976-5895	Sequence 5895, Ap
35	216	5.0	92	4	US-09-621-976-5901	Sequence 5901, Ap
36	216	5.0	92	4	US-09-621-976-5909	Sequence 5909, Ap
37	216	5.0	92	4	US-09-621-976-5914	Sequence 5914, Ap
38	212.5	4.9	776	4	US-10-020-079-24	Sequence 24, Appl
39	212.5	4.9	863	4	US-10-020-079-32	Sequence 32, Appl
40	212.5	4.9	876	4	US-10-020-079-30	Sequence 30, Appl
41	212.5	4.9	895	4	US-10-020-079-18	Sequence 18, Appl
42	212.5	4.9	976	4	US-10-020-079-28	Sequence 28, Appl
43	212.5	4.9	982	4	US-10-020-079-26	Sequence 26, Appl
44	210.5	4.9	83	4	US-09-621-976-5396	Sequence 5396, Ap
45	208.5	4.8	789	4	US-10-020-079-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1

US-09-949-016-9182
; Sequence 9182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9182
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9182

Alignment Scores:
Pred. No.: 1.03e-84 Length: 157
Score: 771.00 Matches: 155
Percent Similarity: 97.48% Conservative: 0
Best Local Similarity: 97.48% Mismatches: 2
Query Match: 17.62% Indels: 17
DB: 4 Gaps: 0

US-09-974-546C-85 (1-2506) x US-09-949-016-9182 (1-157)

QY 31 TGTATAATAATTACAAAATTTCTTCTATTGCTACAGCTACAAATTTACAGT 90

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

QM nucleic - protein search, using frame_plus_n2p model
Run on: September 9, 2005, 12:21:48 ; Search time 141.556 Seconds
(without alignments)
13965.459 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gacctaataatagaggt.....agtaaaactccccccccc 2506

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1777461 seqs, 394431504 residues
Total number of hits satisfying chosen parameters: 3554922

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09974546/runat_07092005_174400_19612/app_query.fasta_1.4942
-DB=Published Applications AA -QFMT=fastan -SURFIX=n2p.rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=spect -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09974546@cgn_1_229_0runat_07092005_174400_19612
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp:
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp:
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp:
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp:
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp:
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp:
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp:
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp:
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp:
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp:
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp:
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp:
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp:
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp:
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp:
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pcp:
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pcp:
18: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pcp:
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pcp:
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pcp:
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp:
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	694	15.9	135	10	US-09-974-546-84
2	694	15.9	135	10	US-09-974-546-86
3	301.5	7.0	109	15	US-10-121-016-48
4	295.5	6.8	126	14	US-10-082-830-207
5	293.5	6.8	341	18	US-10-220-335-586
6	290.5	6.7	97	15	US-10-112-944-908
7	284.5	6.6	170	14	US-10-017-161-1108
8	284.5	6.6	170	15	US-10-292-798-940
9	283.5	6.6	128	15	US-10-276-774-2194
10	283.5	6.6	152	14	US-10-198-070-112
11	283	6.5	127	14	US-10-198-070-25
12	282.5	6.5	119	9	US-09-764-847-696
13	282.5	6.5	119	14	US-10-092-154-696
14	282.5	6.5	127	15	US-10-108-260A-2558
15	282	6.5	129	14	US-10-198-070-23
16	282	6.5	136	15	US-10-104-047-3447
17	281	6.5	120	15	US-10-108-260A-4121
18	280.5	6.5	144	14	US-10-011-585A-145
19	279.5	6.5	134	15	US-10-416-314-45
20	279	6.4	133	15	US-10-276-774-1788
21	278.5	6.4	136	15	US-10-094-749-1763
22	278.5	6.4	310	9	US-09-800-723-190
23	277.5	6.4	127	11	US-09-833-245-1427
24	277.5	6.4	239	9	US-09-800-729-193
25	277.5	6.4	239	14	US-10-198-070-113
26	277.5	6.4	239	18	US-10-756-149-5485
27	275.5	6.4	109	9	US-09-989-919-105
28	275.5	6.4	156	15	US-10-104-047-1984
29	275.5	6.4	170	10	US-09-986-480-282
30	275.5	6.4	170	17	US-10-863-332-282
31	275.5	6.4	198	10	US-09-986-480-433
32	275.5	6.4	198	17	US-10-863-332-433
33	274.5	6.3	100	13	US-10-016-157A-187
34	274.5	6.3	134	15	US-10-276-774-2463
35	274	6.3	167	14	US-10-017-161-1154
36	274	6.3	167	15	US-10-292-798-974
37	273.5	6.3	107	9	US-09-989-920-233
38	273.5	6.3	241	18	US-10-486-977-2
39	272	6.3	122	11	US-09-833-245-1659
40	271	6.3	171	18	US-10-496-905-352
41	270	6.2	190	14	US-10-017-161-1734
42	270	6.2	190	15	US-10-292-798-1390
43	269.5	6.2	116	15	US-10-108-260A-2525
44	269.5	6.2	123	15	US-10-104-047-2112
45	269.5	6.2	125	15	US-10-108-260A-4066

ALIGNMENTS

RESULT 1
US-09-974-546-84
; Sequence 84, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert

TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS, PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE

NUMBER OF SEQUENCES: 87

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION NUMBER: US/09/974,546
/ FILING DATE: 10-Oct-2001
/ CLASSIFICATION: Unknown
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/097,199
/ FILING DATE: 1998-06-12
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 84:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 135 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-974-546-84

Alignment Scores:
Pred. No.: 6,08e-71 Length: 135
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.86% Indels: 0
DB: 10 Gaps: 0

US-09-974-546C-85 (1-2506) x US-09-974-546-84 (1-135)

QY 99 ATGAGGGCTTCTTAAGGAACAGAAATATGAGGATATGACAAATATTATTCACATTTTA 158
Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisLeu 20
QY 159 CAGATCAGAAAATTGAGGCACAGATTAAGTAATCTCCCAAGGCTACAGGCAATTCAGCT 218
Db 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
QY 219 CCAGAAACTGTCTTACCATTCTGCTACAGGTTATTCGAAAAAAGCAAGTAAAA 278
Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysVallys 60
QY 279 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
QY 339 CTCACACCCCTTGACAGACACATTGACCATGAAGTTCTCAATGAATGTTCTTCATTA 398
Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
QY 399 TCTTCAGAGGACATATTATTACATTGCTTTCAGTTAACTTCAGACCCCTAGTCTGGAA 458
Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
QY 459 TGCTGTCTTCTTACTTATCCAAAACATATACATCCACAGATCATTA 503
Db 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

RESULT 2

US-09-974-546-86
/ Sequence 86, Application US/09974546
/ Publication No. US20030050470A1
/ GENERAL INFORMATION:
/ APPLICANT: An, Gang
/ O'Hara, S. Mark
/ Ralph, David

/ Veltri, Robert
/ TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
/ PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
/ NUMBER OF SEQUENCES: 87
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Arnold, White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/974,546
/ FILING DATE: 10-Oct-2001
/ CLASSIFICATION: Unknown
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/097,199
/ FILING DATE: 1998-06-12
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 86:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 135 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

Alignment Scores:
Pred. No.: 6,08e-71 Length: 135
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.86% Indels: 0
DB: 10 Gaps: 0

US-09-974-546C-85 (1-2506) x US-09-974-546-86 (1-135)

QY 99 ATGAGGGCTTCTTAAGGAACAGAAATATGAGGATATGACAAATATTATTCACATTTTA 158
Db 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisLeu 20
QY 159 CAGATCAGAAAATTGAGGCACAGATTAAGTAATCTCCCAAGGCTACAGGCAATTCAGCT 218
Db 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
QY 219 CCAGAAACTGTCTTACCATTCTGCTACAGGTTATTCGAAAAAAGCAAGTAAAA 278
Db 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysVallys 60
QY 279 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
Db 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
QY 339 CTCACACCCCTTGACAGACACATTGACCATGAAGTTCTCAATGAATGTTCTTCATTA 398
Db 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
QY 399 TCTTCAGAGGACATATTATTACATTGCTTTCAGTTAACTTCAGACCCCTAGTCTGGAA 458
Db 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 9, 2005, 09:43:33 ; Search time 30.2749 Seconds
(without alignments)
15928.642 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gacctaataatagaggt.....agtaaaactccccccacc 2506

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DBV=xlh
-Q=/cgn2 1/USPTO.spool/US09974546/runat.07092005.174358.19508/app_query.fasta_1.4942
-DB=pir 79 -OPMT=fastan -SUFFIX=n2p.tpr -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOALIGN=200 -THRT=SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CGN 1 1 79 @runat.07092005.174358.19508 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DBV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	239.5	5.5	613	4 C40201	artifact-warning s
C 2	231.5	5.3	613	4 C40201	artifact-warning s
C 3	228	5.3	597	4 E40201	artifact-warning s
C 4	224	5.1	597	4 E40201	artifact-warning s
C 5	204	4.7	627	4 A40201	artifact-warning s
C 6	194.5	4.4	673	4 F40201	artifact-warning s
C 7	190.5	4.4	673	4 F40201	artifact-warning s
C 8	187.5	4.3	627	4 A40201	artifact-warning s
C 9	173.5	4.0	196	2 I38022	hypothetical prote
C 10	161	3.7	579	4 D40201	artifact-warning s
C 11	149	3.4	418	2 S41044	chromosomal protei
C 12	145	3.4	301	4 B40201	artifact-warning s
C 13	142	3.3	579	4 D40201	artifact-warning s
C 14	141.5	3.3	100	2 A46010	X-linked retinopat

15	133.5	3.1	114	2	JC5238	galactosylceramide
16	132.5	3.0	124	2	A47582	B-cell growth fact
C 17	129	3.0	499	2	S65657	alpha-1C-adrenerg
C 18	120.5	2.8	301	4	B40201	artifact-warning s
C 19	120	2.8	342	2	PC4211	hepatocellular car
C 20	120	2.8	841	1	I78885	serine/threonine-s
C 21	103.5	2.4	381	2	F71196	probable hexosyltr
C 22	103	2.4	1513	2	D90129	hypothetical prote
C 23	100.5	2.3	369	2	A53959	thromboxane A-2 re
C 24	100.5	2.3	619	2	A60646	transforming protein
C 25	98	2.2	1715	2	T30559	resistance protein
C 26	93.5	2.2	79	2	A56194	thromboxane A-2 re
C 27	93.5	2.2	407	2	T02670	probable thromboxa
C 28	92.5	2.1	263	2	A82069	hypothetical prote
C 29	92	2.1	461	1	GQRT11	tumor necrosis fac
C 30	92	2.1	536	2	T37840	probable vacuolar
C 31	90	2.1	2539	2	B71619	hypothetical prote
C 32	89.5	2.0	296	2	S01441	class II histocomp
C 33	89.5	2.0	399	2	D86322	hypothetical prote
C 34	88.5	2.0	252	2	PC4259	ferritin associate
C 35	88.5	2.0	795	2	S30874	hypothetical prote
C 36	87.5	2.0	863	2	C88546	protein R107.4 lim
C 37	87	2.0	379	2	C75006	hypothetical prote
C 38	87	2.0	493	2	S46826	hypothetical prote
C 39	86.5	2.0	451	2	D70045	two-component sens
C 40	86.5	2.0	459	2	S06607	23S rRNA intron 2
C 41	86	2.0	699	2	S51454	EST1 protein - yea
C 42	85.5	2.0	523	2	D64555	conserved hypothet
C 43	85.5	2.0	679	2	B96599	protein F20N2.12 l
C 44	85.5	2.0	1386	1	RNLVC2	DNA-directed RNA p
C 45	85.5	2.0	3119	2	T18414	protein g377 - mal

ALIGNMENTS

RESULT 1
C40201
artifact-warning sequence (translated ALU class C) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: C40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: C40201
A:Molecule type: DNA
A:Residues: 1-613 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A:Reference number: A40200; MUID:92241891; PMID:1572661
C:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of thi

Alignment Scores:	3.88e-17	Length:	613
Pred. No.:	239.50	Matches:	61
Score:	73.63%	Conservative:	6
Percent Similarity:	67.03%	Mismatches:	17
Best Local Similarity:	5.54%	Indels:	8
Query Match:	4	Gaps:	2
DB:			
US-09-974-546C-85 (1-2506) x C40201 (1-613)			
QY	2483	GTTCGCCAGCTGGAGTGCATCTAGCTACCGCAACACCTCACCTCC	2424
Db	322	ValThrGlnAlaGlyValGlnTrpArgAspLeuGlySerLeuGlnAlaProPro---	340
QY	2423	GGGTTTCACGCAATTCCTCGC-CTCAGCCTCCCTGAGTAGCTGGGNTTACAGGATGCGCA	2365

```
Db 341 GlyPheMetProPheSerCysLeuSerLeuLeuArgThrTrpAspTyrArgArgPro-Hi 360
QY 2364 CCACGCCCGCTAATTTTGTGTTTATAGATGGGGTTCCACCATGCTGTCAG 2305
Db 360 SHisAlaGlnLeuIlePheCysIlePheSerArgAsn-GlyValLeuProCysTrpProG 380
QY 2304 GCTGGTCTCGAACTCCGACCTCAGTGATCCACCA-----CCTCCAAAGTGTGGG 2251
Db 380 lYTrpSerArgThrProGlyLeu--MetIleHisProLeuProLeuProProLysValLeuGl 399
QY 2250 ATTACAGGCGTGAGCCACCGTGCCT 2226
Db 399 yLeuGlnAla**AlaThrAlaPro 407

RESULT 2
C40201
artifact-warning sequence (translated ALU class C) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: C40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: C40201
A:Molecule type: DNA
A:Residues: 1-613 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A:Reference number: A40200; PMID:1572661
C:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

Alignment Scores:
Pred. No.: 2,91e-16 Length: 613
Score: 231.50 Matches: 53
Percent Similarity: 63.33% Conservative: 4
Best Local Similarity: 58.89% Mismatches: 27
Query Match: 5.29% Indels: 6
DB: 4 Gaps: 2

US-09-974-546C-85 (1-2506) x C40201 (1-613)

QY 2227 GGCAGGTGGCTCAGCCCTGTAATCCAGCACTTTGGGA-----GGTGGTGGATCACC 2280
Db 208 GlyAlaValAlaHisAlaCysAsnProSerThrLeuGlyGlySerGlyTrpIle--M 227
QY 2281 TGAGGTGCGGAGTTTCGAGACCCAGCCTGACCATGTTGGAAACCCCATCTCTACTATAAA 2340
Db 227 eArgProGlyValArgAspGlnProGlyGlnHisGlyLysThrProPhe-LeuLeuLys 246
QY 2341 ACACAAAAAATTAGCGCGGTGGTGGCCATCCCTGTATCCAGCTACTCAGGAGGCT 2400
Db 247 IleGlnLysIleSerTrpAlaTrpCysGlyArgLeu**SerHisValArgArgLeu 266
QY 2401 GAGGCGGAGAAATGCTTGTAACCCGAGAGTGGAGTGTGGTGGATGAGTGCATGCAC 2460
Db 267 A-rGlnGlnAsnGlyIleAsnPro--GlyGlyGlyAlaCysSerGluProArgSerArg 285
QY 2461 TATTGCACCTCCAGCCTGGGCAACA 2484
Db 286 HisCysThrProAlaTrpValThr 293

RESULT 3
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
```

```
personal communication, 1992
A:Reference number: A40201
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A:Reference number: A40200; PMID:1572661
C:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of thi

Alignment Scores:
Pred. No.: 6,99e-16 Length: 597
Score: 228.00 Matches: 65
Percent Similarity: 72.45% Conservative: 6
Best Local Similarity: 66.33% Mismatches: 16
Query Match: 5.27% Indels: 13
DB: 4 Gaps: 4

US-09-974-546C-85 (1-2506) x E40201 (1-597)

QY 2495 AGTTTACTCTGTTGCCAGCTGGAGTGCATAGTCAATCTCAGCTCACC-----GCA 2440
Db 507 SerPheAlaLeuValAlaGlnAlaGlyValArgTrpHisAsnLeuThrAlaAsnPheAla 526
QY 2439 ACAACCTCCACTCCCGGTTCAAGCAATTCGCGC---CTCAGCCTCCTCAGTAGCTGG 2383
Db 527 SerTrp-----ValGlnAlaIleLeuSerCysLeuSerLeuProSerSerTrp 542
QY 2382 GATTACAGCATGTGCCACACCGCGCTAATTTTGTGTTTGTAGTAGATGGGG 2323
Db 543 AspTyrArgHisAlaProProArgProAlaAsn--PheIlePheLeuValGluMet-Gly 561
QY 2322 TTCCACCATGCTGGTCAGGCTGCTCGAACTCCGACCTCAGGTGATCCACCCAC----- 2267
Db 562 PheLeuHisValGlyGlnAlaGlyLeuLysLeuProThrSerGlyAspProProArgLeu 581
QY 2266 ---CTCCAAAGTGTGGGATTACAGGCGTGAGCCACCGTCGCTGGCC 2222
Db 582 GluLeuProLysArgTrpAspTyrArgHisGluLeuProHisLeuAla 597

RESULT 4
E40201
artifact-warning sequence (translated ALU class E) - human
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence_revision 11-Aug-1995 #text_change 19-May-2000
C:Accession: E40201
R:Claverie, J.M.
personal communication, 1992
A:Reference number: A40201
A:Accession: E40201
A:Molecule type: DNA
A:Residues: 1-597 <CLA>
R:Claverie, J.M.
Genomics 12, 838-841, 1992
A:Title: Identifying coding exons by similarity search: Alu-derived and other potentially
A:Reference number: A40200; PMID:1572661
C:Contents: annotation
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of or
in-frame stop codons are shown as 'X'.
C:Comment: Any significant similarity of a predicted protein sequence to a portion of thi

Alignment Scores:
Pred. No.: 1,92e-15 Length: 597
Score: 224.00 Matches: 60
Percent Similarity: 68.04% Conservative: 6
Best Local Similarity: 61.86% Mismatches: 23
Query Match: 5.12% Indels: 9
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: September 9, 2005, 06:29:13 ; Search time 149.193 Seconds
(without alignments)
17202.865 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 4375
Sequence: 1 gacctaataatcgaggt.....agtaaaactccccccccc 2506

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09974546/runat.07092005.174358.19494/app.query.fasta.1.4942
-DB=UniProt.03 -QFMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CGN.1.1.302 @runat.07092005.174358.19494 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt.03:*
1: uniprot_sprot:*
2: uniprot_treml:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	15.9	135	Q9GZY1	Q9GZY1 homo sapien
2	333.5	7.6	593	ALU6 HUMAN	P31933 homo sapien
3	327	7.6	593	ALU6 HUMAN	P31933 homo sapien
4	316.5	7.2	593	ALU7_HUMAN	P31914 homo sapien
5	313.5	7.2	129	Q6ZSV6	Q6ZSV6 homo sapien
6	308	7.1	593	ALU7_HUMAN	P31914 homo sapien
7	297.5	6.9	140	Q6ZP99	Q6ZP99 homo sapien
8	288.5	6.7	129	Q6ZUW8	Q6ZUW8 homo sapien
9	286.5	6.6	164	Q6ZUK0	Q6ZUK0 homo sapien
10	283.5	6.6	152	Q6ZK85	Q6ZK85 homo sapien
11	283	6.5	127	Q9H743	Q9H743 homo sapien
12	282	6.5	129	Q9H9H0	Q9H9H0 homo sapien
13	281.5	6.5	125	Q6ZTE1	Q6ZTE1 homo sapien
14	281.5	6.5	260	Q6ZQR8	Q6ZQR8 homo sapien
15	281	6.5	120	Q8N8A3	Q8N8A3 homo sapien
16	280	6.5	132	Q6ZNU7	Q6ZNU7 homo sapien

c 17	278.5	6.4	134	2	Q6ZSG8	Q6ZSG8 homo sapien
c 18	278.5	6.4	136	2	Q6NR6	Q6NR6 homo sapien
c 19	278.5	6.4	146	2	Q6ZTX8	Q6ZTX8 homo sapien
c 20	277.5	6.4	239	2	Q9NX17	Q9NX17 homo sapien
c 21	277	6.3	591	1	ALU8_HUMAN	P31915 homo sapien
c 22	276	6.4	132	2	Q6ZS28	Q6ZS28 homo sapien
c 23	276	6.4	591	1	ALU8_HUMAN	P31915 homo sapien
c 24	275.5	6.3	131	2	Q6ZNV7	Q6ZNV7 homo sapien
c 25	275.5	6.4	156	2	Q8NBH4	Q8NBH4 homo sapien
c 26	274.5	6.3	142	2	Q6ZWD5	Q6ZWD5 homo sapien
c 27	271.5	6.3	239	2	Q6ZWA9	Q6ZWA9 homo sapien
c 28	269.5	6.2	125	2	Q8N8C2	Q8N8C2 homo sapien
c 29	269	6.2	148	2	Q6ZUC5	Q6ZUC5 homo sapien
c 30	268	6.2	118	2	Q9H387	Q9H387 homo sapien
c 31	267	6.2	202	2	Q6ZUA4	Q6ZUA4 homo sapien
c 32	266.5	6.2	130	2	Q6ZSV2	Q6ZSV2 homo sapien
c 33	266	6.1	123	2	Q9HAD8	Q9HAD8 homo sapien
c 34	264	6.1	180	2	Q8N7M7	Q8N7M7 homo sapien
c 35	263.5	6.1	122	2	Q6ZNZ6	Q6ZNZ6 homo sapien
c 36	263.5	6.1	128	2	Q6ZPB2	Q6ZPB2 homo sapien
c 37	260.5	6.0	123	2	Q9PIN7	Q9PIN7 homo sapien
c 38	260.5	6.0	146	2	Q6ZRX6	Q6ZRX6 homo sapien
c 39	260.5	6.0	152	2	Q6ZUG4	Q6ZUG4 homo sapien
c 40	260	6.0	124	2	Q6ZUN5	Q6ZUN5 homo sapien
c 41	260	6.0	585	1	ALU5_HUMAN	P31912 homo sapien
c 42	258.5	6.0	142	2	Q6ZRQ0	Q6ZRQ0 homo sapien
c 43	258.5	6.0	179	2	Q8N1K7	Q8N1K7 homo sapien
c 44	257.5	6.0	167	2	Q6ZSN3	Q6ZSN3 homo sapien
c 45	257.5	6.0	174	2	Q8N2A0	Q8N2A0 homo sapien

ALIGNMENTS

RESULT 1
Q9GZY1 PRELIMINARY; PRT; 135 AA.
ID Q9GZY1
AC Q9GZY1
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE UC28 protein.
GN Name=UC28;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate cancer;
RX MEDLINE=21028101; PubMed=11156405;
RA An G., Ng A.Y., Meka C.S.R., Luo G., Bright S.P., Cazares L.,
RT Wright G.L. Jr., Veltri R.W.;
RT "Cloning and characterization of UROC28, a novel gene overexpressed in
RT prostate, breast, and bladder cancers.";
RL Cancer Res. 60:7014-7020(2000).
DR EMBL; AF189270; AAG17118.1; -;
DR EMBL; AF189269; AAG17117.1; -;
DR Genew; HGNC:21079; PROVL.
SQ SEQUENCE 135 AA; 15722 MW; 2B7DB8B1983705D0 CRC64;

Alignment Scores:
Pred. No.: 4.25e-63 Length: 135
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.86% Indels: 0
DB: 2 Gaps: 0

US-09-974-546C-85 (1-2506) x Q9GZY1 (1-135)

Qy 99 ATGAGCGCTCTTAAAGCAACAGAAATATGAGGATATGCACATATTATTACATTTTA 158
|||||
Db 1 MetArgAlaPheLeuArgAsnGlnIlyrGluAspMetHisAsnIleIleLeu 20
|||||

QY 159 CAGATCAGAAATTGAGCAGACAGATTAGTAATCTCCCAAGGCTACCAAGCATTCTAGCT 218
 Db |||||
 21 GlnIleArgLysLeuArgHisArgLeuSerAsnProArgLeuProGlyIleLeuAla 40
 QY 219 CAGAAACTGCTCTTACCATTCGTCTACAGAGTATTCGAAAAAAGAAAGTAAAA 278
 Db |||||
 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
 QY 279 AGAAGTCAAAAGCAACAGATTGATTATTCATAGACAGTCACACCATGCAATT 338
 Db |||||
 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 QY 339 CTCACACCCCTTCGAGACACATTCGACCATGAAGTTCTCAATGAATGTTCTCTCATTA 398
 Db |||||
 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 QY 399 TCTTCAGAACCATATATTACATTCATGCTTTCGAGTTAACTCAGACCCCTAGGTCGGAA 458
 Db |||||
 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 QY 459 TCGTGTCTCTCTACTTATCCAAAACATATACATCCACAGATCATA 503
 Db |||||
 121 CysCysLeuLeuThrLeuSerLysThrIleHisProGlnIlelle 135

RESULT 2

ALU6 HUMAN
 ID ALU6 HUMAN STANDARD; PRT; 593 AA.
 AC P39193;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Alu subfamily SP sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95021758; PubMed=7935834; DOI=10.1038/371752a0;
 RA Claverie J.-M., Makalowski W.,
 RT "Alu alert."
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT.
 RX MEDLINE=92241891; PubMed=1572661;
 RA Claverie J.-M.,
 RT "Identifying coding exons by similarity search: alu-derived and other
 potentially misleading protein sequences."
 RL Genomics 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.,
 RT "The Alu family developed through successive waves of fixation closely
 connected with primate lineage history."
 RL J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J., Milosavljevic A.,
 RT "Reconstruction and analysis of human Alu genes."
 RL J. Mol. Evol. 32:105-121(1991).
 CC -1- MISCELLANEOUS: Various analyses indicate that Alu repeats fall
 into 8 subfamilies (Ref.3 and Ref.4). Therefore, 8 Alu warning
 consensus sequences have been constituted that contain all six
 frames conceptual translations of each of these classes of Alu
 repeats.
 CC -1- MISCELLANEOUS: Isolated 'X' indicates the presence of a stop
 codon. 'XXX' is used to separate the various translation phases.
 CC -1- CAUTION: This Alu entry is provided in order to avoid the further
 pollution of protein sequence databases with Alu-derived amino
 acid sequences.

CC -1- CAUTION: Alu repetitive sequences are interspersed in human and
 primate genomes with an average spacing of 4 kb. Some of them are
 actively transcribed by pol III. Normal transcripts may contain
 Alu-derived sequences in 5' or 3' untranslated regions. However,
 cDNA libraries also contain partial and/or rearranged cDNAs
 ligated with Alu-derived sequence in any orientation. Although Alu
 elements (especially situated on the complementary strand) have a
 great potential to create additional/alternative exons,
 consideration should be given to the possibility that the presence
 of an Alu in an open reading frame may have resulted from a
 cloning artifact or may be due to misinterpretation of sequencing
 data. This point has been overlooked on several occasions, with
 the consequence of erroneous Alu-derived amino acid sequences
 being reported.
 CC -1- CAUTION: Any significant similarity of a putative protein sequence
 with an Alu-translated entry must be taken as a warning that a
 part of Alu repeat may have been artifactually included in the
 coding nucleotide sequence.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U14572; -; NOT ANNOTATED_CDS.
 DR ProDom; PD003738; GIDA; 1.
 KW Hypothetical protein.
 FT DOMAIN 1 97 Frame-1.
 FT DOMAIN 101 196 Frame-2.
 FT DOMAIN 200 295 Frame-3.
 FT DOMAIN 299 395 Frame-4.
 FT DOMAIN 399 494 Frame-5.
 FT DOMAIN 498 593 Frame-6.
 SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;
 Alignment Scores:
 Pred. No.: 2,9e-25 Length: 593
 Score: 333.50 Matches: 76
 Percent Similarity: 84.38% Conservative: 5
 Best Local Similarity: 79.17% Mismatches: 10
 Query Match: 7.62% Indels: 6
 DB: 1 Gaps: 2
 US-09-974-546C-85 (1-2506) x ALU6_HUMAN (1-593)
 QY 2222 GGCCAGGACGGTGGCTCACGCTGTATCCAGCACTTTGG-----GAGTGGGTGA 2275
 Db 1 GlyArgAlaArgTrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlyGly 20
 QY 2276 TCACCTGAGTGGGAGTTCGAGACCCAGCTGACCATGGTGAACCCCATCTCTAC 2335
 Db 21 SerProGluValGlySerArgProAla***ProThrTrpArgAsn-ProValSerTh 40
 QY 2336 TAAAAACACAAAAAATTAGCCGGGTGGTGGACATCGCTGTATCCAGCTACTCAGG 2395
 Db 40 LysAsnThr-LysIleSerArgAlaTrpTrpArgMetProValIleProAlaThrArgG 60
 QY 2396 AGGCTGAGGC-GGAGAAATGCTTGAACCCGGAAGGTGGAGTGTGTCGGTGGAGTGA 2454
 Db 60 LuAlaGluAlaGlyGluSerLeuGluProGlyArg---ArgArgLeuArg***AlaGluI 79
 QY 2455 TTGCACTATTGCACTCCAGCTGGGCAACAGAGGTAAAACTCCC 2498
 Db 79 leAlaProLeuHisSerSerLeuGlyAsnLysSerGluThrPro 93
 RESULT 3
 ALU6 HUMAN
 ID ALU6 HUMAN STANDARD; PRT; 593 AA.
 AC P39193;
 DT 01-FEB-1995 (Rel. 31, Created)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 8, 2005, 22:24:09 ; Search time 7220.71 Seconds
(without alignments)
16816.746 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gaccttaatatatcgaggt.....agtaaaactccccccacccc 2506

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2505.6	100.0	2512	9	AF189269 Homo sapi
2	2494	99.5	2505	6	AR146836 Sequence
3	2424.2	96.7	129227	9	HS171N11 Human DNA
4	2424.2	96.7	186698	2	AC036236 Homo sapi
5	2417	96.4	2496	9	BC069109 Homo sapi
6	1961.6	78.3	2103	9	AF189270 Homo sapi
7	1950	77.8	2087	6	AR146835 Sequence
8	626.2	25.0	757	6	AR146757 Sequence
9	201.6	8.0	137227	9	AC092933 Homo sapi
10	201.6	8.0	166895	2	AC023415 Homo sapi
11	201.6	8.0	170961	2	AC112202 Homo sapi
12	201.6	8.0	174054	2	AC131148 Homo sapi
13	198.8	7.9	200237	9	AF168787 Homo sapi
14	198.6	7.9	129625	9	AC009301 Homo sapi
15	198.6	7.9	318586	2	AC107422 Homo sapi
16	195.6	7.8	73499	9	AC132942 Homo sapi
17	195.6	7.8	209885	2	AC011189 Homo sapi
18	194.4	7.8	120551	9	AL391215 Human DNA
19	194.2	7.7	156886	2	AC023445 Homo sapi

C	20	194.2	7.7	174551	9	AC010598	Homo sapi
C	21	193	7.7	56884	2	AC104588	Homo sapi
C	22	192.8	7.7	164107	9	AC104982	Homo sapi
C	23	192.4	7.7	150636	2	AC027529	Homo sapi
C	24	192.4	7.7	192203	9	AC034102	Homo sapi
C	25	192.2	7.7	146957	9	AC084264	Homo sapi
C	26	192.2	7.7	180346	9	AC092615	Homo sapi
C	27	191.8	7.7	131637	9	CNS01DUC	Human chr
C	28	191.8	7.7	164514	9	CNS01RGV	Human chr
C	29	191.2	7.6	110000	2	BX324168_0	Homo sapi
C	30	191.2	7.6	112884	9	AL136966	Human DNA
C	31	191.2	7.6	162377	9	AL683870	Human DNA
C	32	191	7.6	187490	2	AC048378	Homo sapi
C	33	190.6	7.6	109588	9	AC010461	Homo sapi
C	34	190.6	7.6	161920	9	AC025169	Homo sapi
C	35	190.4	7.6	148984	9	HS105D16	Human DNA
C	36	190.2	7.6	182945	2	AC069004	Homo sapi
C	37	190.2	7.6	210643	9	AC137723	Homo sapi
C	38	190	7.6	128379	9	HSJ773A18	Human DNA
C	39	190	7.6	163521	2	AC108040	Homo sapi
C	40	190	7.6	196773	9	AC114480	Homo sapi
C	41	189.8	7.6	201939	9	AC044792	Homo sapi
C	42	189.6	7.6	170502	9	AC093573	Pan trogl
C	43	189.4	7.6	147859	2	AC011079	Homo sapi
C	44	189.4	7.6	157289	2	AC021447	Homo sapi
C	45	189.4	7.6	178305	2	AC139475	Homo sapi

ALIGNMENTS

RESULT 1
AF189269
LOCUS Homo sapiens UC28 protein (UC28) mRNA, complete cds.
DEFINITION
ACCESSION AF189269
VERSION AF189269.1 GI:10441601
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS An.G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L., Wright,G.L. Jr. and Veltri,R.W.
TITLE Cloning and characterization of UROC28, a novel gene overexpressed in prostate, breast, and bladder cancers
JOURNAL Cancer Res. 60 (24), 7014-7020 (2000)
MEDLINE 21028101
PUBMED 11156405
REFERENCE 2 (bases 1 to 2512)
AUTHORS An.G. and Veltri,R.W.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research Parkway, Oklahoma City, OK 73104, USA

FEATURES
source
1. .2512
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q23-q24"
/tissue_type="prostate cancer"
1. .2512
/gene="UC28"
93.506
/gene="UC28"
/codon_start=1
/product="UC28 protein"
/protein_id="AAG17117.1"
/db_xref="GI:10441602"
/translation="MRAFLRNQKYMNMNIHLQIRKLRHLSNFPRLPGILAPETV
LPPFCYKVRKKRKSQKATEFIDYIEQSHHALLTPIQTHLTWKGSMSKSSLS


```
Db 2041 |||||ACAGTACTGATACATAATTTAAAGGCACACTCCCTTCCACTTTGGTAAACACAGACTCT 2100
Qy 2101 AATTGGACCAACCTGAGCTTAGGACTACAGGCATACAAATAGTAACTCTGTCACGA 2160
Db 2101 AATTGGACCAACCTGAGCTTAGGACTACAGGCATACAAATAGTAACTCTGTCACGA 2160
Qy 2161 TTCACCTCATCTGTGATTTCTCTAGATGTTTACTAGGCGTTGTTATATAAATAACCC 2220
Db 2161 TTCACCTCATCTGTGATTTCTCTAGATGTTTACTAGGCGTTGTTATATAAATAACCC 2220
Qy 2221 CGGCCAGGCAGCTGGCTCAGCGCTGTAAATCCAGCACCTTTGGGAGGTGGGTGGAATCACC 2280
Db 2221 CGGCCAGGCAGCTGGCTCAGCGCTGTAAATCCAGCACCTTTGGGAGGTGGGTGGAATCACC 2280
Qy 2281 TGAGTCCGGAGTTTCAGACACGCTGACAGCATGCTGTAATCCAGCTACTCAGGAGGCT 2340
Db 2281 TGAGTCCGGAGTTTCAGACACGCTGACAGCATGCTGTAATCCAGCTACTCAGGAGGCT 2340
Qy 2341 ACACAAATAATAGCCGGCGTGGTGGCACATGCTGTAATCCAGCTACTCAGGAGGCT 2400
Db 2341 ACACAAATAATAGCCGGCGTGGTGGCACATGCTGTAATCCAGCTACTCAGGAGGCT 2400
Qy 2401 GAGCGCGAGAAATGCTTTGAACCCGGAAGGTGGAGGTTGTTGCGTGCAGCTGAGATTGCAC 2460
Db 2401 GAGCGCGAGAAATGCTTTGAACCCGGAAGGTGGAGGTTGTTGCGTGCAGCTGAGATTGCAC 2460
Qy 2461 TATTGCACCTCAGCCCTGGGCAACAGGAGTAAACTCCCCCCCCACC 2506
Db 2461 TATTGCACCTCAGCCCTGGGCAACAGGAGTAAACTCCCCCCCCACC 2506

RESULT 2
ARI146836
LOCUS      Sequence 85 from patent US 6218529.
DEFINITION
ACCESSION  ARI146836
VERSION    ARI146836.1 GI:15110025
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 2505)
AUTHORS   An,G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
TITLE     Biomarkers and targets for diagnosis, prognosis and management of
          prostate, breast and bladder cancer
JOURNAL   Patent: US 6218529-A 85 17-APR-2001;
FEATURES   Location/Qualifiers
           source 1..2505
              /organism="unknown"
              /mol_type="unassigned DNA"

ORIGIN
Query Match 99.5%; Score 2494; DB 6; Length 2505;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2505; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 GACCTTAAATATATCAGGTGGCTTAATTGATGATATAATAATTTACAAAATTTATCTCTA 60
Db 1 GACCTTAAATATATCAGGTGGCTTAATTGATGATATAATAATTTACAAAATTTATCTCTA 60
Qy 61 TTGCTACAGAGCTACAATTCAAATTTACAGTAGGCCACCATGAGGCGCTTTTAAAGGAACC 120
Db 61 TTGCTACAGAGCTACAATTCAAATTTACAGTAGGCCACCATGAGGCGCTTTTAAAGGAACC 120
Qy 121 AGAAATATGAGGATATGCACAAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
Db 121 AGAAATATGAGGATATGCACAAATATTATTCACATTTTACAGATCAGAAAATTGAGGCACA 180
Qy 181 GATTAAAGTAACTTCCCAAGGCTACCAGGCAATTCAGCTCCAGAAACTGTGCTCTTACCAT 240
Db 181 GATTAAAGTAACTTCCCAAGGCTACCAGGCAATTCAGCTCCAGAAACTGTGCTCTTACCAT 240
```

```
Qy 241 TCTGCTACAAGGTATTTTCGAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGACT 300
Db 241 TCTGCTACAAGGTATTTTCGAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGACT 300
Qy 301 TCATTTGATTATTTCCATAGAACAGTACACCATGCAATTTCTCACACCTTTGCAGACACT 360
Db 301 TCATTTGATTATTTCCATAGAACAGTACACCATGCAATTTCTCACACCTTTGCAGACACT 360
Qy 361 TGACCATGAAAGGTTCTCAATGAAATGTTCTCTCAATTTATCTTCAGAGCCATATTATTC 420
Db 361 TGACCATGAAAGGTTCTCAATGAAATGTTCTCTCAATTTATCTTCAGAGCCATATTATTC 420
Qy 421 CAATTGACCTTTGAGTTAACTCAGAGCCCTAGGCTGGAATGCTGCTCTCTACTTATCCA 480
Db 421 CAATTGACCTTTGAGTTAACTCAGAGCCCTAGGCTGGAATGCTGCTCTCTACTTATCCA 480
Qy 481 AAAATATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA 540
Db 481 AAAATATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA 540
Qy 541 ATAAAAATGGTTGAAAGGCAATTTCTGCTAACATGCACTGTTTAAAGCCCGCCAGTAAC 600
Db 541 ATAAAAATGGTTGAAAGGCAATTTCTGCTAACATGCACTGTTTAAAGCCCGCCAGTAAC 600
Qy 601 TGAACCATTTCCAACTTCAATTTTACATGAAAGAAATTTGATGATCTAGGAGGTTATTT 660
Db 601 TGAACCATTTCCAACTTCAATTTTACATGAAAGAAATTTGATGATCTAGGAGGTTATTT 660
Qy 661 AATTCCTAAAAATACAAACCCATCTGATCTTCTCAATCTTGAACTCATAGATTATTTCT 720
Db 661 AATTCCTAAAAATACAAACCCATCTGATCTTCTCAATCTTGAACTCATAGATTATTTCT 720
Qy 721 ATTATCTCAATTTAGTTTGTATTTATCTAGTGGGCCATTAAAACTACCACTATGTGT 780
Db 721 ATTATCTCAATTTAGTTTGTATTTATCTAGTGGGCCATTAAAACTACCACTATGTGT 780
Qy 781 TCTGCTCTCCATTTAGTCAATTAACCTAACTAAGCAATTTAGTAAAGCCATGTCAGAT 840
Db 781 TCTGCTCTCCATTTAGTCAATTAACCTAACTAAGCAATTTAGTAAAGCCATGTCAGAT 840
Qy 841 GCTCCGCTAGGCACAGAGGGATAAAAACTATTATAGTATACCACTAAATTTTCGCTT 900
Db 841 GCTCCGCTAGGCACAGAGGGATAAAAACTATTATAGTATACCACTAAATTTTCGCTT 900
Qy 901 AGTAACTAGTGAATTTCAAGTCACTGCTGAGTCAAGAGTTTGAGGAGACATTACAATGT 960
Db 901 AGTAACTAGTGAATTTCAAGTCACTGCTGAGTCAAGAGTTTGAGGAGACATTACAATGT 960
Qy 961 GTAATGGAAACCAAGGAAAGTGAATTTGATTAAGTGGGAGCTAGTGTATTATATT 1020
Db 961 GTAATGGAAACCAAGGAAAGTGAATTTGATTAAGTGGGAGCTAGTGTATTATATT 1020
Qy 1021 TAAATTTGATTTCTGACTCTATCATTTGGCCCTCCAAACACAGATTGTGTTTTCTTGGTTTT 1080
Db 1021 TAAATTTGATTTCTGACTCTATCATTTGGCCCTCCAAACACAGATTGTGTTTTCTTGGTTTT 1080
Qy 1081 GTTTTCTTCACTATGGGATCTTCTGTCGCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Db 1081 GTTTTCTTCACTATGGGATCTTCTGTCGCCAGCACAGTGCCTGACACATAGAAAACAATC 1140
Qy 1141 AATATTGCTGAATTAATGATTAATAAATFAGAGAACTTTCCCATCTGTTTGGATCTAT 1200
Db 1141 AATATTGCTGAATTAATGATTAATAAATFAGAGAACTTTCCCATCTGTTTGGATCTAT 1200
Qy 1201 AGAATCATCCAGAGTAAAGTGAAGGCGCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
Db 1201 AGAATCATCCAGAGTAAAGTGAAGGCGCTCTGCATTTATATGCGCTTAAATTAAGATTAT 1260
Qy 1261 GTGAGAAAGTTTAAAGACACTTACTAGTGTGATTTGAAAATATAGTAAACACTTTGGAAA 1320
Db 1261 GTGAGAAAGTTTAAAGACACTTACTAGTGTGATTTGAAAATATAGTAAACACTTTGGAAA 1320
Qy 1321 TGGTGGTGCTTTTAAAAAGATATTAATAGATATATGAAAAATCTCCATCTCAAAAAATAATG 1380
```

Db 1321 TGTGTGCTTTAAAGGATATTAATAGATAATGANAATCTCCATCTCAAAAAATAATG 1380
QY CATAAACTATTTAAAGGAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
Db CATAACTATTTAAAGGAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTTT 1440
QY CATATATTTTACCATTCTGCTGAAGGAGTATATCAAAAGGTTAAAGAAAGATGGAGGA 1500
Db CATATATTTTACCATTCTGCTGAAGGAGTATATCAAAAGGTTAAAGAAAGATGGAGGA 1500
QY AAACCTAGTAAGAAATTTATATTAGTCTGTTTGCAGGTTAGAAAGATTTCTCATCTCAA 1560
Db AAACCTAGTAAGAAATTTATATTAGTCTGTTTGCAGGTTAGAAAGATTTCTCATCTCAA 1560
QY CTTTATGACGAGGAGGAGGCTGTTTGAAGACCATTTTACTTAGCAGAACCAATAT 1620
Db CTTTATGACGAGGAGGAGGAGGCTGTTTGAAGACCATTTTACTTAGCAGAACCAATAT 1620
QY TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAACTTTGTTGATCAAC 1680
Db TTTAGACACTTCCCTGCATTAACTGCACAAACAATATGTTTGCAACTTTGTTGATCAAC 1679
QY CTCACAAACGACATTCAGGAGTTAAATATTTTTCATCAACATTTGATTTTCCTTA 1740
Db CTCACAAACGACATTCAGGAGTTAAATATTTTTCATCAACATTTGATTTTCCTTA 1739
QY AGCTAGAGATTGCTACAAATCTTCTGAAGGCTCAATGGCTTCAAGGCTTGAAGAGAT 1800
Db AGCTAGAGATTGCTACAAATCTTCTGAAGGCTCAATGGCTTCAAGGCTTGAAGAGAT 1799
QY TTTCCCTCTGTTAAGCAGACAGACAAATTAGCCATTTTCACTCTCAAACTTCACTAATGA 1860
Db TTTCCCTCTGTTAAGCAGACAGACAAATTAGCCATTTTCACTCTCAAACTTCACTAATGA 1859
QY TCACATCTTTTCAAAGGAACTTAGAAGACAAATGCCCCGAGTTAAGAACATCAAAA 1920
Db TCACATCTTTTCAAAGGAACTTAGAAGACAAATGCCCCGAGTTAAGAACATCAAAA 1919
QY CTAACCATCTGAAGAACTTTCCCAAGTGAAGCTGCGCATTTTCACTCTCAAACTTCACTAATGA 1980
Db CTAACCATCTGAAGAACTTTCCCAAGTGAAGCTGCGCATTTTCACTCTCAAACTTCACTAATGA 1979
QY GGACTCAAAACAGTCTTTCTCCTGTTGCTGTTTCTGCTCCAGACCAAGGCACTGAG 2040
Db GGACTCAAAACAGTCTTTCTCCTGTTGCTGTTTCTGCTCCAGACCAAGGCACTGAG 2039
QY ACAGTACTGATACATAATTTAAAGCACACTCCCTTCCACTTTGGTAAATACCAGAACTCT 2100
Db ACAGTACTGATACATAATTTAAAGCACACTCCCTTCCACTTTGGTAAATACCAGAACTCT 2099
QY AATTGGACCACTGAAGCTTAGGACTACCGCATACAAATAGTAAACTCTGTCCAGA 2160
Db AATTGGACCACTGAAGCTTAGGACTACCGCATACAAATAGTAAACTCTGTCCAGA 2159
QY TTCACTCATCTGTGTTATTTCTATAGATGTTTACTAGGCGTTGTTTATATAAAATACCC 2220
Db TTCACTCATCTGTGTTATTTCTATAGATGTTTACTAGGCGTTGTTTATATAAAATACCC 2219
QY CGGCCAGGACCGGTGCTCAACGCTGTAATCCAGACATTTGGGAGGTGGGTGGATCACC 2280
Db CGGCCAGGACCGGTGCTCAACGCTGTAATCCAGACATTTGGGAGGTGGGTGGATCACC 2279
QY TCAGGTTCGGAGTTTCAGACACAGCTGACACATGTTGGACACCCCATCTCTACTAAA 2340
Db TCAGGTTCGGAGTTTCAGACACAGCTGACACATGTTGGACACCCCATCTCTACTAAA 2339
QY ACACAAAAAATTAGCGGCGGTGGGACACATGCTGTAATCCAGCTACTCAGGAGGT 2400
Db ACACAAAAAATTAGCGGCGGTGGGACACATGCTGTAATCCAGCTACTCAGGAGGT 2399
QY GAGGCGGAGAAATTTGTTGAACCGGAGGTGAGGTTGTTGGGTGAGCTGAGATTGCAC 2460
Db GAGGCGGAGAAATTTGTTGAACCGGAGGTGAGGTTGTTGGGTGAGCTGAGATTGCAC

Db 2400 GAGGCGGAGAAATTTGTTGAACCGGAGGTGAGGTTGTTGGGTGAGCTGAGATTGCAC 2459
QY 2461 TATTTCACCTCCAGCTCGGCAACAGAGTAAAACTCCCCCACC 2506
Db 2460 TATTTCACCTCCAGCTCGGCAACAGAGTAAAACTCCCCCACC 2505
RESULT 3
HS171N11 129227 bp DNA linear PRI 04-MAR-2003
LOCUS Human DNA sequence from clone RP1-171N11 on chromosome 6q23,
DEFINITION complete sequence.
ACCESSION AL031433
VERSION AL031433.4 GI:4826442
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 129227)
AUTHORS Williams,S.
TITLE Direct Submission
JOURNAL Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 13, 1999 this sequence version replaced gi:4582115.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP1-171N11 is from the library RPCI-1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
Location/Qualifiers
1..129227
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="q23"
/clone_lib="RP1-171N11"
/clone_lib="RPCI-1"

FEATURES
source
1..129227
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="q23"
/clone_lib="RP1-171N11"
/clone_lib="RPCI-1"

ORIGIN
Query Match 96.7%; Score 2424.2; DB 9; Length 129227;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2494; Conservative 1; Mismatches 4; Indels 8; Gaps 6;

QY 1 GACCTTAAATATATCGAGGTGGCTAATTGATGATATTAATTAATTTACAAAATTTCTTCTTA 60
Db 29796 GACCTTAAATATATCGAGGTGGCTAATTGATGATATTAATTAATTTACAAAATTTCTTCTTA 29855
QY 61 TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCATGAGGGCTCTTTAAGGAAC 120
Db 29856 -TGCTACAGAGCTACAAATTCAAATTTACAGTA-GTAAATATGAGGGCTCTTTAAGGAAC 29913
QY 121 AGAAATATGAGGATATGCACAAATTAATTAATTTACAGATTCAGAAATTTGAGGCACA 180
Db 29914 AGAAATATGAGGATATGCACAAATTAATTTACAGATTCAGAAATTTGAGGCACA 29973
QY 181 GATTAAGTAACTTCCCAAGGCTACAGGCAATTCAGTCCAGAAATGTCGCTTTACCAT 240
Db 29974 GATTAAGTAACTTCCCAAGGCTACAGGCAATTCAGTCCAGAAATGTCGCTTTACCAT 30033
QY 241 TCTGCTACAGGTATTTTCGAAAAAGAAAAAGTAAAAAGAGTCAAAAAGGCAACAGAGT 300
Db 30034 TCTGCTACAGGTATTTTCGAAAAAGAAAAAGTAAAAAGAGTCAAAAAGGCAACAGAGT 30093
QY 301 TCATGATTAATCCATAGAACAGTCACACATGCAATTCACACCCCTTGACAGACACAT 360
Db 30094 TCATGATTAATCCATAGAACAGTCACACATGCAATTCACACCCCTTGACAGACACAT 30153
QY 361 TGACCATGAAGGTTCTCAATGAATGTTCCCTCATTAATTCAGAAAGCCATATTATCA 420
Db 30154 TGACCATGAAGGTTCTCAATGAATGTTCCCTCATTAATTCAGAAAGCCATATTATCA 30213
QY 421 CATTTGACTTTGCAAGTTAACTCAGACCCCTAGTCTGGAATGCTCTCTCTACTATATCA 480
Db 30214 CATTTGACTTTGCAAGTTAACTCAGACCCCTAGTCTGGAATGCTCTCTCTACTATATCA 30273
QY 481 AAATATATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAAAGCCCTTCCAGAAA 540
Db 30274 AAATATATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAAAGCCCTTCCAGAAA 30333
QY 541 ATAAAAATGTTGAAAAGGCAATTTCTGTACCAATGACTGTTTAAAGCCGCAAGTAAC 600
Db 30334 ATAAAAATGTTGAAAAGGCAATTTCTGTACCAATGACTGTTTAAAGCCGCAAGTAAC 30393
QY 601 TGAACCAATTCCAACTTCAATTTACTTATGAAGAATTTGATGATGTAGGAGTTATTTTC 660
Db 30394 TGAACCAATTCCAACTTCAATTTACTTATGAAGAATTTGATGATGTAGGAGTTATTTTC 30453
QY 661 AATTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTAATATCT 720
Db 30454 AATTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTAATATCT 30513
QY 721 ATTAATCTCAATTTAGTTTGTATTTATTCCTAGTGGGCCAATTAATAAATCTACCAATGTGT 780
Db 30514 ATTAATCTCAATTTAGTTTGTATTTATTCCTAGTGGGCCAATTAATAAATCTACCAATGTGT 30573
QY 781 TCTGTCTCTCCATTAATTAATCAATTAACAGCAATTTAGTAAAGCCATGTGCCAGAT 840
Db 30574 TCTGTCTCTCCATTAATTAATCAATTAACAGCAATTTAGTAAAGCCATGTGCCAGAT 30633
QY 841 GCTCCGCTAGGCACACAGAGGATAAAAAATATCTTATAGTATACCACTAAATTTTCGCTT 900
Db 30634 GCTCCGCTAGGCACACAGAGGATAAAAAATATCTTATAGTATACCACTAAATTTTCGCTT 30693
QY 901 AGTAACTAGTGAATGTTCAAGTCATGCTGAGTCAGAGTTGAGGAGCAATTTACAAATGT 960
Db 30694 AGTAACTAGTGAATGTTCAAGTCATGCTGAGTCAGAGTTGAGGAGCAATTTACAAATGT 30753
QY 961 GTAATGGAACCAAGGAAGTGAACCTTTGGATAAGTGGGACTAGTGTATTTATATTT 1020
Db 30754 GTAATGGAACCAAGGAAGTGAACCTTTGGATAAGTGGGACTAGTGTATTTATATTT 30813
QY 1021 TAATTGATTTCTGACTCTATCTATCTGCTCCAAACACAGATCTGTTTCTTTGTTGTTT 1080
Db 30814 TAATTGATTTCTGACTCTATCTATCTGCTCCAAACACAGATCTGTTTCTTTGTTGTTT 30873

QY 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCACAGTGCCTGCACATAGAAAAACAATC 1140
Db 30874 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCACAGTGCCTGCACATAGAAAAACAATC 30933
QY 1141 -AATATTTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCAATCTGTTTGGATCTA 1199
Db 30934 AATATTTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCAATCTGTTTGGATCTA 30993
QY 1200 TAGAACATCCAGAGTAAGTGTAGGGGCTCTGCATTTATATGCGCTTAAATTAAGATTA 1259
Db 30994 TAGAACATCCAGAGTAAGTGTAGGGGCTCTGCATTTATATGCGCTTAAATTAAGATTA 31053
QY 1260 TGTGAGAAAAAGTTTAAAGACACTTAGTAGAGTGAATTTTGAATAATAGTAAACACTTGGAA 1319
Db 31054 TGTGAGAAAAAGTTTAAAGACACTTAGTAGAGTGAATTTTGAATAATAGTAAACACTTGGAA 31113
QY 1320 ATGCTGGTGTCTTTAAAAAGATTAATTAAGATAATAGAAATCTCCATCTCAAAAATAAT 1379
Db 31114 ATGCTGGTGTCTTTAAAAAGATTAATTAAGATAATAGAAATCTCCATCTCAAAAATAAT 31173
QY 1380 GCATAAACATTTTAAAGGAAAAATCAGTCTCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
Db 31174 GCATAAACATTTTAAAGGAAAAATCAGTCTCAGGCTTTCAATGTTTGTTCATTACTTTT 31233
QY 1440 TCATATATTTTACCATCTCTGTAAGGAGTCAATCAAGGGTAAAGAAAGATGGGAG 1499
Db 31234 TCATATATTTTACCATCTCTGTAAGGAGTCAATCAAGGGTAAAGAAAGATGGGAG 31293
QY 1500 AAAAATCTCAGTAAGAAATTAATTAAGTCTGTTGCAAAAGTAGAAAAAGATTTCTCATCTCA 1559
Db 31294 AAAAATCTCAGTAAGAAATTAATTAAGTCTGTTGCAAAAGTAGAAAAAGATTTCTCATCTCA 31353
QY 1560 ACCTTATGAGCAGGAGGAGGAGGCTGTTTGAGAACCAATTTATAGCAGAACCAACATA 1619
Db 31354 ACCTTATGAGCAGGAGGAGGAGGCTGTTTGAGAACCAATTTATAGCAGAACCAACATA 31413
QY 1620 TTTTGAACACTTCCCTGCATTAACCTGCACAAACATATGTTTGAACACTTGTTCATCAA 1679
Db 31414 TTTTGAACACTTCCCTGCATTAACCTGCACAAACATATGTTTGAACACTTGTTCATCAA 31473
QY 1680 CCTCAACAAACGACACATTTAGGAGTTAAATATTTTTCATCAAAATTTGGAATTTTCTTT 1739
Db 31474 CCTCAACAAACGACACATTTAGGAGTTAAATATTTTTCATCAAAATTTGGAATTTTCTTT 31533
QY 1740 AACGCTAGAGATTGCTACAAATCTTCTGAAGGTTCTAAATGGCTTCAGCTAAGAGAGA 1799
Db 31534 AACGCTAGAGATTGCTACAAATCTTCTGAAGGTTCTAAATGGCTTCAGCTAAGAGAGA 31593
QY 1800 TTTTCTCCCTGTTTATAGCAGCAAGCAAAATAGCCATTTCACTCTCAAACTTCACTAATG 1859
Db 31594 TTTTCTCCCTGTTTATAGCAGCAAGCAAAATAGCCATTTCACTCTCAAACTTCACTAATG 31653
QY 1860 ATCACAATCTTTTCAAAAAGGAATCTAGAAAGCAAAATGCCCCGAGTTAAGAAACATCAAA 1919
Db 31654 ATCACAATCTTTTCAAAAAGGAATCTAGAAAGCAAAATGCCCCGAGTTAAGAAACATCAAA 31713
QY 1920 ACTAACCATCTGAAGAACTTCCCAAGTGTAGAGCTCTGCCAATTAATAAATTCAGAGAG 1979
Db 31714 ACTAACCATCTGAAGAACTTCCCAAGTGTAGAGCTCTGCCAATTAATAAATTCAGAGAG 31773
QY 1980 GGGACTCAACAGTCTTTCTTCTTGTGTTGTTTCTGCTCCAGACCAAGGCACTGAC 2039
Db 31774 GGGACTCAACAGTCTTTCTTCTTGTGTTGTTTCTGCTCCAGACCAAGGCACTGAC 31833
QY 2040 GACAGTACTGATACATAATTTAAAAAGCACACTTCCCTGCTCCAGACCAAGGCACTGAC 2099
Db 31834 GACAGTACTGATACATAATTTAAAAAGCACACTTCCCTGCTCCAGACCAAGGCACTGAC 31892
QY 2100 TAATTGGACCACTCTGAAGCTTTAGGACTACAGCCATCAAAATAGTAACTCTGTCCAG 2159
Db 31893 TAATTGGACCACTCTGAAGCTTTAGGACTACAGCCATCAAAATAGTAACTCTGTCCAG 31951
QY 2160 ATTCACTCAATCTGTGTTATTTCTTAGATGTTTACTAGCGGTTTGTATATATAAATAATACC 2219


```
|||||
Db 31952 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTATATAAAATACC 32011
QY 2220 CGGGCCAGGACGGTGGCTCAGCCCTGTAATCCAGCACTTTGGGAGGGTGGTGCATCAC 2279
Db 32012 CGGGCCAGGACGGTGGCTCAGCCCTGTAATCCAGCACTTTGGGAGGGTGGTGCATCAC 32071
QY 2280 CTGAGTCCGGGAGTTCCGAGACAGGCTGACACGACGTGGTGGAAACCCCACTCTCTACTAAA 2339
Db 32072 CTGAGTCCGGGAGTTCCGAGACAGGCTGACACGACGTGGTGGAAACCCCACTCTCTACTAAA 32131
QY 2340 AACACAAAAATAGCCGGCGTGGTGGCACATGCTGTAATCCAGCACTTACGAGGCG 2399
Db 32132 AACACAAAAATAGCCGGCGTGGTGGCACATGCTGTAATCCAGCACTTACGAGGCG 32191
QY 2400 TCAGGGCGGAGATTCCTTGAACCCGGAAGGTGGAGTTTGTGGGTGAGCTGAGATTGCA 2459
Db 32192 TCAGGGCGGAGATTCCTTGAACCCGGAAGGTGGAG---GTTGGGTGAGCTGAGATTGCA 32248
QY 2460 CTATTGCACTCCAGCTGGGCAACAGGAGTAAACTCCCCCCCACC 2506
Db 32249 CTATTGCACTCCAGCTGGGCAACAGGAGTAAACTCCCCCCCACC 32295

RESULT 4
AC036236 186698 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 6 clone RP11-133015, WORKING DRAFT
DEFINITION AC036236 SEQUENCE, 7 unordered pieces.
ACCESSION AC036236
VERSION AC036236.2 GI:9958166
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston.R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 186698)
Waterston.R.H.
Direct Submission
Submitted (07-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:7523932.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0133015
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator; plasmid; 0%
Chemistry: Dye-terminator; Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 183373 bases at least Q40
Consensus quality: 184432 bases at least Q30
Consensus quality: 185064 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 186098; sum-of-contigs
Quality coverage: 5.84 in Q20 bases; agarose-fp
Quality coverage: 5.57 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
```

```
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4769: contig of 4769 bp in length
* 4869: gap of unknown length
* 4770 4869: contig of 6788 bp in length
* 4870 11657: contig of 6788 bp in length
* 11658 11757: gap of unknown length
* 11758 28189: contig of 16432 bp in length
* 28190 28289: gap of unknown length
* 28290 53441: contig of 25052 bp in length
* 53442 102327: gap of unknown length
* 102328 102427: gap of 48886 bp in length
* 102428 182781: contig of 80354 bp in length
* 182782 182882: gap of unknown length
* 182882 186698: contig of 3817 bp in length.
FEATURES
source
1..186698
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clones="RP11-133015"
1..4769
/note="assembly_name:Contig10"
4870..11657
/note="assembly_name:Contig11"
11758..28189
/note="assembly_name:Contig12"
28290..53341
/note="assembly_name:Contig13"
53442..102327
/note="assembly_name:Contig14"
102428..182781
/note="assembly_name:Contig15"
clone_end:SP6
vector_end:right
182882..186698
/note="assembly_name:Contig9"

Query Match 96.7%; Score 2424.2; DB 2; Length 186698;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2494; Conservative 1; Mismatches 4; Indels 8; Gaps 6;
QY 1 GACCTTAATATATCGAGTGGCTAAATGATGATATAATTTACAAATATTCTTCTTA 60
Db 159156 GACCTTAATATATCGAGTGGCTAAATGATGATATAATTTACAAATATTCTTCTTA 159215
QY 61 TTGCTACAGAGCTACAAATTTCAATTTACAGTAGGCCACCATGAGGCGCTTTCTTAAGGAACC 120
Db 159216 -TGCTACAGAGCTACAAATTTCAATTTACAGTA-GTAAATATGAGGCGCTTTCTTAAGGAACC 159273
QY 121 AGAAATATGAGGATATGCACAAATATTATTCACATTTTACAGATCAGAAAAATTGAGGCGACA 180
Db 159274 AGAAATATGAGGATATGCACAAATATTATTCACATTTTACAGATCAGAAAAATTGAGGCGACA 159333
QY 181 GATTAAAGTAATCTCCCAAGGCTACAGGCACTTAGCTCCAGAACTGTGCTTTACCAT 240
Db 159334 GATTAAAGTAATCTCCCAAGGCTACAGGCACTTAGCTCCAGAACTGTGCTTTACCAT 159393
QY 241 TCTGCTACAAGGTATTTTCGAAAAAGAAAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 300
Db 159394 TCTGCTACAAGGTATTTTCGAAAAAGAAAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 159453
QY 301 TCATTGATTTTCCATAGAACAGTCACACCATGCAATTTCTCACCCCTTGAGACACACT 360
Db 159454 TCATTGATTTTCCATAGAACAGTCACACCATGCAATTTCTCACCCCTTGAGACACACT 159513
QY 361 TGACCATGAAAGGTTCCCTCAATGAAATGTTCTCTCATTTCTTCAAGAGCCATATTATTCA 420
Db 159514 TGACCATGAAAGGTTCCCTCAATGAAATGTTCTCTCATTTCTTCAAGAGCCATATTATTCA 159573
```


Qy	421	CATTGAC	TTTTCGAGTTAACTCAGACCCCTAGGTC	TGGAATGCTGTCTTCTCTACTATCA	480
Db	159574	CATTGAC	TTTTCGAGTTAACTCAGACCCCTAGGTC	TGGAATGCTGTCTTCTCTACTATCA	159633
Qy	481	AAACTATATAC	TCCACAGATCATATAAACTCTCAGCCCTGCTC	CAAGCCTTTCCAGAAA	540
Db	159634	AAACTATATAC	TCCACAGATCATATAAACTCTCAGCCCTGCTC	CAAGCCTTTCCAGAAA	159693
Qy	541	ATAAAAAATGGT	TGAAAAGCAATTCCTGCTACCAATGACTGTTT	TAAGCCAGCAAGTAAC	600
Db	159694	ATAAAAAATGGT	TGAAAAGCAATTCCTGCTACCAATGACTGTTT	TAAGCCAGCAAGTAAC	159753
Qy	601	TGAACCAATCCAACTTCAATTTATCTTATGAAAGAAATTTGATGATG	TAGGAGGTTATTTTC	660	
Db	159754	TGAACCAATCCAACTTCAATTTATCTTATGAAAGAAATTTGATGATG	TAGGAGGTTATTTTC	159813	
Qy	661	AAATCTAAAAATACAAAACCAATGTTGATCTTCTCAATCTTGAACCTCATAGATTA	TATATCT	720	
Db	159814	AAATCTAAAAATACAAAACCAATGTTGATCTTCTCAATCTTGAACCTCATAGATTA	TATATCT	159873	
Qy	721	ATTATCTCAATTTAGTTTGTATTATTTATCCTAGTGGCCATTA	AAAAACTACCACATGTGTT	780	
Db	159874	ATTATCTCAATTTAGTTTGTATTATTTATCCTAGTGGCCATTA	AAAAACTACCACATGTGTT	159933	
Qy	781	TCTGTCTCTCCATTTAGTCAATAAATAACAGCAATTA	TAGTAAAGCCATGTGCCAGAT	840	
Db	159934	TCTGTCTCTCCATTTAGTCAATAAATAACAGCAATTA	TAGTAAAGCCATGTGCCAGAT	159993	
Qy	841	GCTCGCTAGGACACAGAGGGATAAAACATACATTATAGTATA	CCACTAAATTTTCGCTT	900	
Db	159994	GCTCGCTAGGACACAGAGGGATAAAACATACATTATAGTATA	CCACTAAATTTTCGCTT	160053	
Qy	901	AGTAACTAGTGAATGTTTCAAGTCAATGCTGAGTCAAGAGTTT	GAGGAGACATTAACAATGT	960	
Db	160054	AGTAACTAGTGAATGTTTCAAGTCAATGCTGAGTCAAGAGTTT	GAGGAGACATTAACAATGT	160113	
Qy	961	GTAATGGAAAACAAGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATAT	1020		
Db	160114	GTAATGGAAAACAAGAAAGTGAAACTTTGGATAAGTGGGGACTAGTGTATTTATATAT	160173		
Qy	1021	TAAATGGATTTCTGACTCTCATATGGCCCTCCAAAACACAGATTTGCTTTTTCTTTGGTTTTT	1080		
Db	160174	TAAATGGATTTCTGACTCTCATATGGCCCTCCAAAACACAGATTTGCTTTTTCTTTGGTTTTT	160233		
Qy	1081	GTTTTCTTCACTATGGGATCTTCTGTGCCACAGACAGTGCCTGACACATAGAAAAACAATC	1140		
Db	160234	GTTTTCTTCACTATGGGATCTTCTGTGCCACAGACAGTGCCTGACACATAGAAAAACAATC	160293		
Qy	1141	-AATAATTTGCTGAAATAATGATTA	AAAAAATCAGAGAACTTTCCCATTCGTTTGGATCTA	1199	
Db	160294	AAATAATTTGCTGAAATAATGATTA	AAAAAATCAGAGAACTTTCCCATTCGTTTGGATCTA	160353	
Qy	1200	TAGAACATCCAGAGTTAAGTGATGAGGGCTCTGCAATTTATATGCGCTTAAATTAAGATT	1259		
Db	160354	TAGAACATCCAGAGTTAAGTGATGAGGGCTCTGCAATTTATATGCGCTTAAATTAAGATT	160413		
Qy	1260	TGTGAGAAAAGTTTAAAGACACTT	AGTAGAGTGATTTTGAATATAGTAAACACTTTGGAA	1319	
Db	160414	TGTGAGAAAAGTTTAAAGACACTT	AGTAGAGTGATTTTGAATATAGTAAACACTTTGGAA	160473	
Qy	1320	ATGTTGGTGTCTTTAAAAAGATA	TTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT	1379	
Db	160474	ATGTTGGTGTCTTTAAAAAGATA	TTAATAGATAATATGAAAATCTCCATCTCAAAAATAAT	160533	
Qy	1380	GCATAAATACTTTAAAGGAAATCA	CATCTCAGGCTTCAATGTTTGTTCATTACTTTT	1439	
Db	160534	GCATAAATACTTTAAAGGAAATCA	CATCTCAGGCTTCAATGTTTGTTCATTACTTTT	160593	
Qy	1440	TCATATATTTTTTACCATCTGCTGAAGGACGT	CATATCAAAGGCTTAAGAAAGATGGGAG	1499	
Db	160594	TCATATATTTTTTACCATCTGCTGAAGGACGT	CATATCAAAGGCTTAAGAAAGATGGGAG	160653	
Qy	1500	AAAACTCAGTAAGAAATTA	TATAGTCTGTTTGCAAAAGTAGAAAAAGATTTCTCATCACTCA	1559	

RESULT 5	BC069109	LOCUS	BC069109	2496 bp	mRNA	linear	PRI 23-JUN-2004
DEFINITION	Homo sapiens prostate and breast cancer overexpressed 1, mRNA (cdna clone IMAGE:7216926).						

ACCESSION	BC069109	Db	60	TGCTACAGAGCTACAAATTCATTTTACAGTA-GTAAATATGAGGGCTTCTTTAAGGAACCA	118
VERSION	BC069109.1	Qy	122	GAATATGAGGATATGCACAATATATTTTACATTTTACAGATCAGAAAATTTGAGGCACAG	181
KEYWORDS		Db	119	GAATATGAGGATATGCACAATATATTTTACATTTTACAGATCAGAAAATTTGAGGCACAG	178
SOURCE	Homo sapiens (human)	Qy	182	ATTAAGTAACCTTCCCAAGGCTACACAGGCTTCTAGCTCCAGAAACTGTGCTCTTACCATT	241
ORGANISM	Homo sapiens	Db	179	ATTAAGTAACCTTCCCAAGGCTACACAGGCTTCTAGCTCCAGAAACTGTGCTCTTACCATT	238
REFERENCE	1 (bases 1 to 2496)	Qy	242	CTGTCTACAGGATATTTCCGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGTT	301
AUTHORS	Klausner, R.L., Collins, F.S., Wagner, L.H., Derge, J.G., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Schaeetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McSwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.	Qy	302	CATTGATTTTCCATAGAACAGTACACCATGCAATTTCTCACCCCTTTCGACAGACACTT	361
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	Db	299	CATTGATTTTCCATAGAACAGTACACCATGCAATTTCTCACCCCTTTCGACAGACACTT	358
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 15899-15903 (2002)	Qy	362	GACCATGAAAGGTTCTCTCAATGAAATGTTCTCTCATTTCTTTCAGAAAGCATATTTTTCAC	421
PUBLISHED	12477932	Db	359	GACCATGAAAGGTTCTCTCAATGAAATGTTCTCTCATTTCTTTCAGAAAGCATATTTTTCAC	418
REFERENCE	2 (bases 1 to 2496)	Qy	422	ATTGACTTTGAGTAACTCAGACCCCTAGGTCTGGAATGCTGTCTCTCTACTTATCCAA	481
AUTHORS	Strausberg, R.	Db	419	ATTGACTTTGAGTAACTCAGACCCCTAGGTCTGGAATGCTGTCTCTCTACTTATCCAA	478
TITLE	Submitted (16-APR-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	Qy	482	AACTATACATCCACAGATCATATAAATCTCTCAGCCCTGCTGCAAGCCCTTTCAGAAAAA	541
JOURNAL	NIH-MGC Project URL: http://mgc.nci.nih.gov	Db	479	AACTATACATCCACAGATCATATAAATCTCTCAGCCCTGCTGCAAGCCCTTTCAGAAAAA	538
PUBLISHED	Contact: MGC help desk	Qy	542	TAAAAATGGTTGAAAAAGGCAATTTCTGTACCAATGACTGTTTAAAGCCAGCAAGTAACT	601
REFERENCE	Email: cgapbs-r@mail.nih.gov	Db	539	TAAAAATGGTTGAAAAAGGCAATTTCTGTACTCAATGACTGTTTAAAGCCAGCAAGTAACT	598
AUTHORS	Tissue Procurement: Anup Madan, University of Iowa	Qy	602	GAACCATTTCCAACTTCAATTTACTTATGAAAAGAAATTTGATGATGATGAGAGGTTATTTCA	661
TITLE	CDNA Library Preparation: Anup Madan, University of Iowa	Db	599	GAACCATTTCCAACTTCAATTTACTTATGAAAAGAAATTTGATGATGATGAGAGGTTATTTCA	658
JOURNAL	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Neurogenomics Research Lab, 200 B EMBR, University of Iowa, Iowa City, IA-52242	Qy	662	ATTCTTAAATACAAACCCATGCTGATCTTCTCAATCTTGAACCTCATAGATTTATATCTA	721
REMARK	anup-madan@uiowa.edu	Db	659	ATTCTTAAATACAAACCCATGCTGATCTTCTCAATCTTGAACCTCATAGATTTATATCTA	718
COMMENT	Jessica Fahey, Tim Nelson, Jae Goon Yoon and Anup Madan	Qy	722	TTATCTCAATTTAGTTTGTATTTTATCTAGTGCCCACTTAAAAACTACCACATGTGTTT	781
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov	Db	719	TTATCTCAATTTAGTTTGTATTTTATCTAGTGCCCACTTAAAAACTACCACATGTGTTT	778
	Series: Plate: Row: Column: 0.	Qy	782	CTGTCTCTCCATTTAGTCAATTAACCTAACTAACGAGCAATTTAGTAAAGCCATGTGCCAGATG	841
FEATURES	Location/Qualifiers	Db	779	CTGTCTCTCCATTTAGTCAATTAACCTAACTAACGAGCAATTTAGTAAAGCCATGTGCCAGATG	838
source	1..2496	Qy	842	CTCCGCTAGGCACACAGAGGGATAAAAACTATTTATAGTATACCACTAAATTTTCGCCTTA	901
	/organism="Homo sapiens"	Db	839	CTCCGCTAGGCACACAGAGGGATAAAAACTATTTATAGTATACCACTAAATTTTCGCCTTA	898
	/mol_type="mRNA"	Qy	902	GTAACCTAGTGAATTTTCAAGTCTATGCTGCTGAGTCAAGAGTTTGAGGAGACATTTACAATGTG	961
	/db_xref="taxon:9606"	Db	899	GTAACCTAGTGAATTTTCAAGTCTATGCTGCTGAGTCAAGAGTTTGAGGAGACATTTACAATGTG	958
	/clone="IMAGE:7216926"	Qy	962	TAATGGAAACCAAGGAAAGTGAATTTGGATTAAGTGGGAGCTAGTGTATTTATATATTT	1021
	/tissue_type="Synthetic constructs"	Db	959	TAATGGAAACCAAGGAAAGTGAATTTGGATTAAGTGGGAGCTAGTGTATTTATATATTT	1018
	/clone_lib="NIH_MGC_242"	Qy	1022	AATTGATTTCTGACTCTATCATTTGCGCTCCAAACACAGATTTGTTTTCTTTGGTTTTG	1081
	/lab_host="XL10"	Db	1019	AATTGATTTCTGACTCTATCATTTGCGCTCCAAACACAGATTTGTTTTCTTTGGTTTTG	1078
	/note="Vector: pUC19"	Qy	1082	TTTTTCTTCACTATGGGATCTTCTGTGCCACACAGTGCCTGCACATAGAAAAACAATC-	1140
ORIGIN		Db	1079	TTTTTCTTCACTATGGGATCTTCTGTGCCACACAGTGCCTGCACATAGAAAAACAATCA	1138
	Query Match 96.4%; Score 2417; DB 9; Length 2496;	Qy	1141	AATATTTGCTGAATAAATGATTAATAAATACAGAGAACTTTTCCCATCTGTGTTTGGATCTAT	1200
	Best Local Similarity 99.4%; Pred. No. 0;	Db	1139	AATATTTGCTGAATAAATGATTAATAAATACAGAGAACTTTTCCCATCTGTGTTTGGATCTAT	1198
	Matches 2488; Conservative 1; Mismatches 6; Indels 8; Gaps 6;				
	2 ACCTTAAATATATCGAGTGGCTAATTTGATGTATTAATAATTTACAAAATTTCTTCTAT	Qy	61		
	1 ACCTTAAATATATCGAGTGGCTAATTTGATGTATTAATAATTTACAAAATTTCTTCTA-	Db	59		
	62 TGCTACAGAGCTACAAATTTTACAGTACGCCACCATGAGGGCTTCTTTAAGGAACCA	Qy	121		

QY 1201 AGAACATCCAGAGTAAGTGAAGGGCTCTGCAATTTATATATGCGCTTAATTAAGATTAT 1260
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1199 AGAACATCCAGAGTAAGTGAAGGGCTCTGCAATTTATATATGCGCTTAATTAAGATTAT 1258
QY 1261 GTGAGAAAAAGTTTAAAGACACATTAGTAGAGTGATTTTGAATAATAGTAAACACTTCGAAA 1320
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1259 GTGAGAAAAAGTTTAAAGACACATTAGTAGAGTGATTTTGAATAATAGTAAACACTTCGAAA 1318
QY 1321 TGTGTGCTTTTAAAGATATTAATAGATAATATGAAATCTCCATCTCAAAAATAATG 1380
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1319 TGTGTGCTTTTAAAGATATTAATAGATAATATGAAATCTCCATCTCAAAAATAATG 1378
QY 1381 CATAAACTATTTAAAGAAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTTCTTTT 1440
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1379 CATAACTATTTAAAGAAAAATCACATCTCCAGGCTTTCAATGTTTGTTCATTTCTTTT 1438
QY 1441 CATATATTTTACCATCTGCTGAAGGAGTTCATATCAAAAGGGTAAAGAAAGATGGGAGGA 1500
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1439 CATATATTTTACCATCTGCTGAAGGAGTTCATATCAAAAGGGTAAAGAAAGATGGGAGGA 1498
QY 1501 AACTCAGTAAGATATATATAGTCTGTTTGAAGTGAAGAAAGATTTCTCATCTCAA 1560
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1499 AACTCAGTAAGATATATATAGTCTGTTTGAAGTGAAGAAAGATTTCTCATCTCAA 1558
QY 1561 CCTTATGACGAGGAGGAGGCTTTTGAGAACCATTTACTTAGCAGAACCATAT 1620
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1559 CCTTATGACGAGGAGGAGGCTTTTGAGAACCATTTACTTAGCAGAACCATAT 1618
QY 1621 TTTAGACATCTCCTGCATTAACATGACACAAACATATGTTTGAACATTTGTTGATCAAC 1680
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1619 TTTAGACATCTCCTGCATTAACATGACACAAACATATGTTTGAACATTTGTTGATCAAC 1678
QY 1681 CTCCTCAACGACACATTCAGGAGTTAAATATTTTTCATCAACATTTGATTTTCTTTA 1740
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1679 CTCCTCAACGACACATTCAGGAGTTAAATATTTTTCATCAACATTTGATTTTCTTTA 1738
QY 1741 AGCTAGAGATTTGCTCAAAATCTTGAAGGCTCAATGGCTTCAAGGCTTAAGAGAT 1800
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1739 AGCTAGAGATTTGCTCAAAATCTTGAAGGCTCAATGGCTTCAAGGCTTAAGAGAT 1798
QY 1801 TTCTCCTGTTTAAAGCAGACAAATTTAGGCAATTTTCACTCTCAAACTTCACTAATCA 1860
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1799 TTCTCCTGTTTAAAGCAGACAAATTTAGGCAATTTTCACTCTCAAACTTCACTAATCA 1858
QY 1861 TCACATTTCTTCAAAAGGAACTCTAGAGACAAATGCCCGAGTTAAGAAACATCAAAA 1920
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1859 TCACATTTCTTCAAAAGGAACTCTAGAGACAAATGCCCGAGTTAAGAAACATCAAAA 1918
QY 1921 CTAACCATCTGAAGAACTTCCCAAGTGAAGACTCTGCATTAAGAAACATTAACGAGG 1980
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1919 CTAACCATCTGAAGAACTTCCCAAGTGAAGACTCTGCATTAAGAAACATTAACGAGG 1978
QY 1981 GGAATCAACAGTCTTCTTCTTCTGCTGTTTCTTGTCTCCAGACCAAGGCACTGAGG 2040
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1979 GGAATCAACAGTCTTCTTCTTCTGCTGTTTCTTGTCTCCAGACCAAGGCACTGAGG 2038
QY 2041 ACAGTACTGATATATTTAAAGCACACTCCCTTCCACTTTGTTGTAATACCAAGACTCT 2100
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2039 ACAGTACTGATATATTTAAAGCACACTCCCTTCCACTTTTGTATACCAAGACTCT 2097
QY 2101 AATTGACACCTGAGGCTTAGGACTACAGGCATACAAAATAGTAAATCTGTCCACGA 2160
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2098 AATTGGA-CACCTGAGGCTTAGGACTACAGGCATACAAAATAGTAAATCTGTCCACGA 2156
QY 2161 TTCATCATCTGTATTTTCTATAGATCTTTTACTAGGCTTTGTATATAAATATACC 2220
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2157 TTCATCATCTGTATTTTCTATAGATCTTTTACTAGGCTTTGTATATAAATATACC 2216
QY 2221 CGGCCAGGACGCTGCTCACGCTCTTAATCCAGACATTTTGGAGGCTGGGTGATCACC 2280
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2217 CGGCCAGGACGCTGCTCACGCTCTTAATCCAGACATTTTGGAGGCTGGGTGATCACC 2276

QY 2281 TGAGCTCGGAGCTTCGAGACAGCCTGACCAGCATGTTGGACCCCATCTCTACTAAAA 2340
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2277 TGAGCTCGGAGCTTCGAGACAGCCTGACCAGCATGTTGGACCCCATCTCTACTAAAA 2336
QY 2341 ACACAAAAAATTAGCGGGCGTGGGCACATGCTGTAATCCAGCTACTCAGGAGGCT 2400
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2337 ACACAAAAAATTAGCGGGCGTGGGCACATGCTGTAATCCAGCTACTCAGGAGGCT 2396
QY 2401 GAGGCGGAGAAATGCTTGAACCCGGAAGGTGGAGGTTGTTGGGTGAGCTGAGATTGCAC 2460
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2397 GAGGCGGAGAAATGCTTGAACCCGGAAGGTGGAG--GTTGGGTGAGCTGAGATTGCAC 2453
QY 2461 TATTGCACCTCCAGCTCGGCAACAGAGTAAACTCCGCCCA 2503
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2454 TATTGCACCTCCAGCTCGGCAACAGAGTAAACTCCGCCCA 2496

RESULT 6
AF189270
LOCUS AF189270 2103 bp mRNA linear PRI 22-MAR-2001
DEFINITION Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION AF189270
VERSION AF189270.1 GI:10441603
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2103)
AUTHORS An.G., Ng.A.Y., Meka.C.S., Luo.G., Bright.S.P., Cazares.L.,
Wright.G.L. Jr. and Velti.R.W.
TITLE Cloning and characterization of UROC28, a novel gene overexpressed
in prostate, breast, and bladder cancers
JOURNAL Cancer Res. 60 (24), 7014-7020 (2000)
MEDLINE 21028101
PUBMED 11156405
REFERENCE 2 (bases 1 to 2103)
AUTHORS An.G. and Velti.R.W.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
Parkway, Oklahoma City, OK 73104, USA
LOCATION/Qualifiers
FEATURES
source
1..2103
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="6"
/map="6q23-q24"
/tissue_type="prostate cancer"
1..2103
/gene="UC28"
99..506
/gene="UC28"
/codon_start=1
/product="UC28 protein"
/protein_id="AAG17118.1"
/db_xref="GI:10441604"
/translation="MRAPLNQKVEDMNIHILQIKRLHRLSFPRLPGLAPETY
LLPFCYKVRKKEKVRKSKATEFIDYSIEQSHAILPLQTHLTMKSSMKCSLS
EAILFTLTLQTLQTLGLECCLLYSKTHPQII"
misc_feature
1961..2103
/gene="UC28"
/note="alternative splicing"
2047..2052
/gene="UC28"

polya_signal
2047..2052
/gene="UC28"

ORIGIN
Query Match 78.3%; Score 1961.6; DB 9; Length 2103;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1968; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1 GACCTTAATATATCGAGGTGCTAATGATGATATAATAATTACAAATTTCTTCTA 60
|||||

1	GACCTTAAATATATCGAGGTGGCTAAATTGATGATGATTAATTAATTTTACAAAATTTACTTCTCTA	60
61	TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCACCATGAGGGCTTCTTTAAGAAACC	120
61	TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCACCATGAGGGCTTCTTTAAGAAACC	120
121	AGAAATATGAGGATATGCACAATATTTATTCACATTTTACAGATCAGAAAATTTGAGGCACA	180
121	AGAAATATGAGGATATGCACAATATTTATTCACATTTTACAGATCAGAAAATTTGAGGCACA	180
181	GATTTAAGTAATCTTCCCAGGCTACCGAGCAATTCAGCTCCAGAAAATGTGTCTCTTACCAT	240
181	GATTTAAGTAATCTTCCCAGGCTACCGAGCAATTCAGCTCCAGAAAATGTGTCTCTTACCAT	240
241	TCTGCTACAAAGGTATTTTCGAAAAAGAAAAAGTAAAAAGAGTCAAAAAGCCAAACAGAGT	300
241	TCTGCTACAAAGGTATTTTCGAAAAAGAAAAAGTAAAAAGAGTCAAAAAGCCAAACAGAGT	300
301	TCATTTGATTAATTCATAGAAACAGTCAACAACATGCAATTTCTCACACCTTTGCAGACACACT	360
301	TCATTTGATTAATTCATAGAAACAGTCAACAACATGCAATTTCTCACACCTTTGCAGACACACT	360
361	TGACCATGAAAAGTTTCTCAATGAAATGTTTCTCATTTATCTTCAGAGGCCATATTATTTCA	420
361	TGACCATGAAAAGTTTCTCAATGAAATGTTTCTCATTTATCTTCAGAGGCCATATTATTTCA	420
421	CATTGACCTTTGTCAGTTAACTCAGAGCCCTAGTGTCTGGAATGCTGTCTCTCTACTTATCCA	480
421	CATTGACCTTTGTCAGTTAACTCAGAGCCCTAGTGTGGAATGCTGTCTCTCTACTTATCCA	480
481	AAACTATACATCCAAGATCATATAAACTCTCAGCCCTGCTGCAAAAGCCTTTCCAGAAAA	540
481	AAACTATACATCCAAGATCATATAAACTCTCAGCCCTGCTGCAAAAGCCTTTCCAGAAAA	540
541	ATAAAAATGGTTGAAAAGGCAATTCCTGCTACCAATGACTGTTTAAAGCCCAAGCAAGTAAC	600
541	ATAAAAATGGTTGAAAAGGCAATTCCTGCTACCAATGACTGTTTAAAGCCCAAGCAAGTAAC	600
601	TGAACCAATCCAACTTCAATTTACTTATGAAAAGAAATTTGATGATGATGAGAGGTATTTTC	660
601	TGAACCAATTCOAATTCOAATTTACTTATGAAAAGAAATTTGATGATGATGAGAGGTATTTTC	660
661	AAATCTAAAAATACAAACCCATGTTGATCTTTTCTCAATCTTTGAACTCATAGATTTATCTCT	720
661	AAATCTAAAAATACAAACCCATGTTGATCTTTTCTCAATCTTTGAACTCATAGATTTATCTCT	720
721	ATTATCTCAATTTAGTTTGGTTATTTATCTCTAGTGGGCCATTTAAAACTACCATGTGTT	780
721	ATTATCTCAATTTAGTTTGGTTATTTATCTCTAGTGGGCCATTTAAAACTACCATGTGTT	780
781	TCTGTCTCTCCATTTAGTCAATACTAACTAACTAAAGCAATTTAGTAAAGCCATGTGCCAGAT	840
781	TCTGTCTCTCCATTTAGTCAATACTAACTAAAGCAATTTAGTAAAGCCATGTGCCAGAT	840
841	GCTCCGCTTAGGCACAGAGGGATAAAAACAATCTTATAGTATACCACTAAATTTTTCGCTT	900
841	GCTCCGCTTAGGCACAGAGGGATAAAAACAATCTTATAGTATACCACTAAATTTTTCGCTT	900
901	AGTAACTAGTGAAATGTTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTTACAAATGT	960
901	AGTAACTAGTGAAATGTTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTTACAAATGT	960
961	GTAATGGAACCAAGGAAGTGAAACTTTTGGATAGTGGGGACCTAGTGTATTTATATATT	1020
961	GTAATGGAACCAAGGAAGTGAAACTTTTGGATAGTGGGGACCTAGTGTATTTATATATT	1020
1021	TAAATTGATTTCTGACTCTATCATTTGGGCTCCAAACACAGATTTGTTTTTTCTTTGGTTTT	1080
1021	TAAATTGATTTCTGACTCTATCATTTGGGCTCCAAACACAGATTTGTTTTTTCTTTGGTTTT	1080
1081	GTTTTCTTCACTATGGGATCTTTCTGTGCCAGACAGTGCCTGACACATAGAAAAACAATC	1140
1081	GTTTTCTTCACTATGGGATCTTTCTGTGCCAGACAGTGCCTGACACATAGAAAAACAATC	1140

Qy	1141	AAATATTTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCATTTCTGTTTGGATCTAT	1200
Db	1141	AAATATTTGCTGAATAAATGATTAATAAATCAGAGAACTTTCCCATTTCTGTTTGGATCTAT	1200
Qy	1201	AGAACATCCAGATGAAGTGAAGGGCCCTCGCATTTATATGCGCTTAAATTAAGATTAT	1260
Db	1201	AGAACATCCAGATGAAGTGAAGGGCCCTCGCATTTATATGCGCTTAAATTAAGATTAT	1260
Qy	1261	GTGAGAAAAGTTTAAAGACACTTGTAGTAGAGTGAATTTTGAAAATATAGTAAACACTTGGAAA	1320
Db	1261	GTGAGAAAAGTTTAAAGACACTTGTAGTAGAGTGAATTTTGAAAATATAGTAAACACTTGGAAA	1320
Qy	1321	TGTTGGTGTCTTTAAAAGAGATTAATTAATAGATAATGAAAATCTCCATCTCAAAAAATAATG	1380
Db	1321	TGTTGGTGTCTTTAAAAGAGATTAATTAATAGATAATGAAAATCTCCATCTCAAAAAATAATG	1380
Qy	1381	CATTAACATATTTAAAGGAAAATCACATCTCCAGGCTTCAATGTTTTGTTTCATTTACTTTTT	1440
Db	1381	CATTAACATATTTAAAGGAAAATCACATCTCCAGGCTTCAATGTTTTGTTTCATTTACTTTTT	1440
Qy	1441	CATATATTTTTTACCATCTGCTGAAGCGAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA	1500
Db	1441	CATATATTTTTTACCATCTGCTGAAGCGAGTCATATCAAAGGGTAAAGAAAGATGGGAGGA	1500
Qy	1501	AAACTCAGTAAGAATTAATATTAGTCTGTTTTGCAAAAGTAGAAAAAGATTTCTCATCACTCAA	1560
Db	1501	AAACTCAGTAAGAATTAATATTAGTCTGTTTTGCAAAAGTAGAAAAAGATTTCTCATCACTCAA	1560
Qy	1561	CTTTATGAGCAGAGAGGGAAGGCTGTTTGAGAACATTTTACTTTAGCAGAACCATAT	1620
Db	1561	CTTTATGAGCAGAGAGGGAAGGCTGTTTGAGAACATTTTACTTTAGCAGAACCATAT	1620
Qy	1621	TTTATAGACATTCCTTCATTAATGTCACAAACAATATGTTTGCAAACTGTTTGTGATCAAC	1680
Db	1621	TTTATAGACATTCCTTCATTAATGTCACAAACAATATGTTTGCAAACTGTTTGTGATCAAC	1680
Qy	1681	CTCCAAACAGCACACATTCAGGAGTTAAATATATTTTTTCATCAAACTTTGGATTTTTCTTTA	1740
Db	1681	CTCCAAACAGCACACATTCAGGAGTTAAATATATTTTTTCATCAAACTTTGGATTTTTCTTTA	1740
Qy	1741	ACGCTAGAGATTCGTCACAAATCTTTGAAGGGTCTCAATGGCTTCAGGCTAAGAAAGAT	1800
Db	1741	ACGCTAGAGATTCGTCACAAATCTTTGAAGGGTCTCAATGGCTTCAGGCTAAGAAAGAT	1800
Qy	1801	TTCTCCCTGTTTAAAGCAGCAGACAAATTTAGCCATTTCACTCTCAAACTTCACTAATGA	1860
Db	1801	TTCTCCCTGTTTAAAGCAGCAGACAAATTTAGCCATTTCACTCTCAAACTTCACTAATGA	1860
Qy	1861	TCATATTTCTTTCAAAGGAACCTTAGAAGACCAAAATGCCCGAGTTAAGAAACATCAAAA	1920
Db	1861	TCATATTTCTTTCAAAGGAACCTTAGAAGACCAAAATGCCCGAGTTAAGAAACATCAAAA	1920
Qy	1921	CTAACCATCTGAAGAACTTTCCCAAGTGTAAAGTCTGCCATTTAAACATTAACCGAGA	1978
Db	1921	CTAACCATCTGAAGAACTTTCCCAAGTGTAAAGTCTGCCATTTAAACATTAACCGAGA	1978
RESULT 7			
AR146835			
LOCUS			
DEFINITION	AR146835	2087 bp	DNA linear PAT 08-AUG-2001
ACCESSION	AR146835	Sequence 83 from patent US 6218529.	
VERSION	AR146835.1	GI:151110024	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

FEATURES		Location/Qualifiers	
source	1..2087	/organism="unknown"	
		/mol_type="unassigned DNA"	
ORIGIN			
Query Match	77.8%;	Score 1950;	DB 6; Length 2087;
Best Local Similarity	99.4%;	Pred. No. 0;	
Matches 1967;	Conservative 0;	Mismatches 10;	Indels 1; Gaps 1;
Qy	1	GACCTTAAATATATCGAGGTGGCTAAATTTGATGTATATAATTTACAAAATTTATCTTCTTA	60
Db	1	GACCTTAAATATATCGAGGTGGCTAAATTTGATGTATATAATTTACAAAATTTATCTTCTTA	60
Qy	61	TTGCTACAGAGCTACAAATTC	120
Db	61	TTGCTACAGAGCTACAAATTC	120
Qy	121	AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTTGAGGCACA	180
Db	121	AGAAATATGAGGATATGCACAATATTATTCACATTTTACAGATCAGAAAATTTGAGGCACA	180
Qy	181	GATTAAAGTAACTTCCCAAGGCTACACAGGCATTTCTAGCTCCAGAAAATGTGCTCTTACCAT	240
Db	181	GATTAAAGTAACTTCCCAAGGCTACACAGGCATTTCTAGCTCCAGAAAATGTGCTCTTACCAT	240
Qy	241	TCTGCTACAGGTATTTTCGAAAAAAGAAAGTAAAGAAAGTCAAAAGGCCAACAGAGT	300
Db	241	TCTGCTACAGGTATTTTCGAAAAAAGAAAGTAAAGAAAGTCAAAAGGCCAACAGAGT	300
Qy	301	TCATTGATATTCCATAGAACAGTCAACCATGCAATTTCTCACACCTTTCGACAGACACT	360
Db	301	TCATTGATATTCCATAGAACAGTCAACCATGCAATTTCTCACACCTTTCGACAGACACT	360
Qy	361	TGACCATGAAAGGTTCTCTCAATGAATGTTCTCATTTATCTTCAGAGGCCATATTATTC	420
Db	361	TGACCATGAAAGGTTCTCTCAATGAATGTTCTCATTTATCTTCAGAGGCCATATTATTC	420
Qy	421	CATTGACTTTGCGAGTTAACTCAGACCCCTAGGCTCGGAATGCTGCTCTCTACTTTATCCA	480
Db	421	CATTGACTTTGCGAGTTAACTCAGACCCCTAGGCTCGGAATGCTGCTCTCTACTTTATCCA	480
Qy	481	AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGGCCCTTTCAGAAAA	540
Db	481	AAACTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGGCCCTTTCAGAAAA	540
Qy	541	ATAAAAATGGTTGAAAGGCAATTCCTGTACCAATGACTGTTTTAAGCCCGAGCAAGTAAC	600
Db	541	ATAAAAATGGTTGAAAGGCAATTCCTGTACCAATGACTGTTTTAAGCCCGAGCAAGTAAC	600
Qy	601	TGAACCATTCCTCAATTTACTTATGAAAAGAAATTTGATGATGTAGGAGGTTATTTC	660
Db	601	TGAACCATTCCTCAATTTACTTATGAAAAGAAATTTGATGATGTAGGAGGTTATTTC	660
Qy	661	AATTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACCTCATAGATTTATTTCT	720
Db	661	AATTCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACCTCATAGATTTATTTCT	720
Qy	721	ATTATCTCAATTTAGTTTGTATTTATCTTAGTGGGCCATTTAAAACTACCACATGTGTT	780
Db	721	ATTATCTCAATTTAGTTTGTATTTATCTTAGTGGGCCATTTAAAACTACCACATGTGTT	780
Qy	781	TCTGTCTCTCATTAGTCAATAAATACTAACTAAGCAATTTAGTAAAGCCATGTGCCAGAT	840
Db	781	TCTGTCTCTCATTAGTCAATAAATACTAAGCAATTTAGTAAAGCCATGTGCCAGAT	840
Qy	841	GCTCCGCTAGGCACAGAGGATAAAAAAATACTTATAGTATATACCACTAAATTTTCGCTT	900
Db	841	GCTCCGCTAGGCACAGAGGATAAAAAAATACTTATAGTATATACCACTAATTTTCGCTT	900
Qy	901	AGTAACTAGTGAATCTTCAAGTCATGCTGAGTCAAGGTTGAGGAGACATTACAACTGT	960
Db	901	AGTAACTAGTGAATCTTCAAGTCATGCTGAGTCAAGGTTGAGGAGACATTACAACTGT	960

RESULT 8

ARI46757

LOCUS

ARI46757 757 bp DNA linear PAT 08-AUG-2001

Qy	961	GTAATGGAACCAAGGAAAGTGAAACTTTTGGATTAAGTGGGACTAGTGTTATTTATATTT	1020
Db	961	GTAATGGAACCAAGGAAAGTGAAACTTTTGGATTAAGTGGGACTAGTGTTATTTATATTT	1020
Qy	1021	TAATTGATTTCTGACTCTATCATTTGSCCTCCAAACACACAGATGTGTGTTTCTTTGGTTTT	1080
Db	1021	TAATTGATTTCTGACTCTATCATTTGSCCTCCAAACACACAGATGTGTGTTTCTTTGGTTTT	1080
Qy	1081	GTTTTCTTCACTATGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAAACAATC	1140
Db	1081	GTTTTCTTCACTATGGATCTTCTGTGCCCAGCACAGTGCCTGACACATAGAAAAACAATC	1140
Qy	1141	ATAATTTGCTGAATAAATGATTAATAAATAATCAGAGAACTTTCCATTTCTGTTGGATCTAT	1200
Db	1141	ATAATTTGCTGAATAAATGATTAATAAATAATCAGAGAACTTTCCATTTCTGTTGGATCTAT	1200
Qy	1201	AGAACATCCAGAGTAAAGTGAAGGCGCTCTGCATTTATATGGCTTAAATTAAGATTAT	1260
Db	1201	AGAACATCCAGAGTAAAGTGAAGGCGCTCTGCATTTATATGGCTTAAATTAAGATTAT	1260
Qy	1261	GTGAGAAAAAGTTTAAAGACACTTAGTAGAGTGAATTTGAAATATATAGTAAACACTTCGAAA	1320
Db	1261	GTGAGAAAAAGTTTAAAGACACTTAGTAGAGTGAATTTGAAATATATAGTAAACACTTCGAAA	1320
Qy	1321	TGGTGGTGTCTTTAAAGATATTAATAGATAATAGAAATCTCCATCTCAAAAAATAATG	1380
Db	1321	TGGTGGTGTCTTTAAAGATATTAATAGATAATAGAAATCTCCATCTCAAAAAATAATG	1380
Qy	1381	CATAAATCTTTTAAAGGAAAAATCAGATCTCCAGGCTTTCATTTGTTGTTTCACTACTTTT	1440
Db	1381	CATAAATCTTTTAAAGGAAAAATCAGATCTCCAGGCTTTCATTTGTTGTTTCACTACTTTT	1440
Qy	1441	CATATATTTTAAACCATCTGCTGAAGCAGTCAATCAAAAGGCTTAAAGAAAGTGGAGGA	1500
Db	1441	CATATATTTTAAACCATCTGCTGAAGCAGTCAATCAAAAGGCTTAAAGAAAGTGGAGGA	1500
Qy	1501	AAATCAGTAAAGATTTATATTAGTCTGTTTGCAGAGTAAAGAAAGATTTCTCATCTCAA	1560
Db	1501	AAATCAGTAAAGATTTATATTAGTCTGTTTGCAGAGTAAAGAAAGATTTCTCATCTCAA	1560
Qy	1561	CCTTATGACGAGAGAGGAGGAGGCTGTTTGAGAACCTTTTACTTAGCAGAACCATAT	1620
Db	1561	CCTTATGACGAGAGAGGAGGAGGCTGTTTGAGAACCTTTTACTTAGCAGAACCATAT	1620
Qy	1621	TTTAGACACTTCCCTGCAATTAAGTGCACAAATATGTTTGCAGAACTTGTGTGATCAAC	1680
Db	1621	TTTAGACACTTCCCTGCAATTAAGTGCACAAATATGTTTGCAGAACTTGTGTGATCAAC	1679
Qy	1681	CTCCAAACACGACACATTCAGGATTAATAATTTTTCATCAAAACATTTGGATTTTCTTTA	1740
Db	1680	CTCCAAACACGACACATTCAGGATTAATAATTTTTCATCAAAACATTTGGATTTTCTTTA	1739
Qy	1741	ACGCTAGAGATTTGCTACAAATCTTCTGAAGGCTCAATGGCTTCAGGCTTAAGAGAGAT	1800
Db	1740	ACGCTAGAGATTTGCTACAAATCTTCTGAAGGCTCAATGGCTTCAGGCTTAAGAGAGAT	1799
Qy	1801	TTTCTCCCTGTTTATAAGCAGACAGAAATTAGCCATTTTCACTCTCAAACTTCACATATGA	1860
Db	1800	TTTCTCCCTGTTTATAAGCAGACAGAAATTAGCCATTTTCACTCTCAAACTTCACATATGA	1859
Qy	1861	TCACATTTCTTTCCAAAAGGAACTCTAGAGACCAAAATGCCGAGTTTAAAGACATCAAAA	1920
Db	1860	TCACATTTCTTTCCAAAAGGAACTCTAGAGACCAAAATGCCGAGTTTAAAGACATCAAAA	1919
Qy	1921	CTAACCATCTGAAGAACTTCCCAAGTGTAGACTCTGCCATTAAAACTATACCGAGA	1978
Db	1920	CTAACCATCTGAAGAACTTCCCAAGTGTAGACTCTGCCATTAAAACTATACCGAGA	1977

DEFINITION Sequence 3 from patent US 6218529.
ACCESSION AR146757
VERSION AR146757.1 GI:15109946
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 757)
AUTHORS An.G., O'Hara,S.Mark., Ralph,D. and Veltri,R.
TITLE Biomarkers and targets for diagnosis, prognosis and management of
prostate, breast and bladder cancer
JOURNAL Patent: US 6218529-A 3 17-APR-2001;
FEATURES Location/Qualifiers
source 1..757
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 25.0%; Score 626.2; DB 6; Length 757;
Best Local Similarity 94.6%; Pred. No. 1.6e-123;
Matches 716; Conservative 0; Mismatches 28; Indels 13; Gaps 6;
QY 1686 ACAACGACATCTCAGGAGTTAAATATTTTTCATCAACATTTGGATTTTTCCTTAACGCT 1745
Db 1 ACAACGACATCTCAGGAGTTAAATATTTTATCATCAACATTTGGATTTTTCCTTAACGCT 60
QY 1746 AGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAAGAGATTTC 1805
Db 61 AGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAAGAGATTTC 120
QY 1806 CTGTGTATAAGCAGCAAGCAAAATTAGCCATTTCACTCTCAAACTTCACATATGATCA 1865
Db 121 CTGTGTATAAGCAGCAAGCAAAATTAGCCATTTCACTCTCAAACTTCACATATGATCA 180
QY 1866 TTCTTTCCAAAGGACTCTAGAGACCAAAATGCCCGATTGAAGACATCAAACTAAC 1925
Db 181 TTCTTTCCAAAGGACTCTAGAGACCAAAATGCCCGATTGAAGACATCAAACTAAC 240
QY 1926 CATCTGAAGAACTTCCCAAGTGTAGACTCTGCCATTAACATTAACGAGGGGACT 1985
Db 241 CATCTGAAGAACTTCCCAAGTGTAGACTCTGCCATTAACATTAACGAGGGGACT 300
QY 1986 CAAACAGTCTTTT - CTTCCTTTGTGCTGTTTC - TTGCTCCCGACCAAGG - ----CACT 2036
Db 301 CAAACAGTCTTTTCTTCTTCCCTTTGTGCTGTTTCTTGTCTCCAGACCAAGGCACTTGGG 360
QY 2037 GAGCAGTACTGATACATAATTTAAAGCA - CACTCCCTTCCACTTTGGTAA - TACCA 2093
Db 361 GAGCAGTACTGATACATAATTTAAAGCAACCACTCCCTTCCCACTTTGTAATACCCA 420
QY 2094 GAACTCTAATTGGACCACTTGAAGCTTAGGA - CTACCAAGCCATACAAATAGTAAACTCT 2152
Db 421 GAACTCTAATTGGACCACTTGAAGCTTAGGACTTACGACCCATACAAATAGTAAACTCT 480
QY 2153 GTCCAGATTCACTCATCTGTGTATTTTCTATAGATGTTTATAGCGTTGTGTATATAA 2212
Db 481 GTCCAGATTCACTCATCTGTGTATTTTCTATAGATGTTTATAGCGTTGTGTATATAA 540
QY 2213 AAATACCCCGGCGAGCGTGGCTCAGCCTGTAAATCCAGCACTTTGGAGGTGGGT 2272
Db 541 AAATACCCCGGCGAGCGTGGCTCAGCCTGTAAATCCAGCACTTTGGAGGTGGGT 600
QY 2273 GGATCACTTGAGTGGGAGTTTCGAGACCAAGCTGACCAAGCATGGTGAACCCCATCTTC 2332
Db 601 GGATCACTTGAGTGGGAGTTTCGAGACCAAGCTGACCAAGCATGGTGAACCCCATCTTC 660
QY 2333 TACTAAAAACAAAAAATTAGCCGGCGTGGTGGCACAATGCTGTAATCCAGCTACTTC 2392
Db 661 TACTAAAAACAAAAAATTAGCCGGCGTGGTGGCACAATGCTGTAATCCAGCTACTTC 720
QY 2393 AGGAGGCTGAGCGGAGNATTCGTTGAACCCGAGG 2429
Db 721 AGGAGGCTGAGCGGAGNATTCGTTGAACCCGAGG 757

RESULT 9
AC092933/c
LOCUS
DEFINITION Homo sapiens 3 BAC RP11-200119 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
AC092933
AC092933.21 GI:22657492
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 137227)
Musny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowles,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Lousheed,H., Lozadro,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nickerson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuben,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williams,A., Wleczkyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 137227)
Worley,K.C.
Direct Submission
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 137227)
Worley,K.C.
Direct Submission
Submitted (30-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 137227)
Worley,K.C.
Direct Submission

Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavskiy, L.,
 Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
 Choepl, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
 DeArellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
 Fencstor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R.,
 Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
 McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,
 Menues, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M.,
 Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
 Travers, M., Trigliio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
 Zody, M.

TITLE

Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 8, 2000 this sequence version replaced gi:7342076.

All repeats were identified using RepeatMasker:

Smit, A. F. A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6745

Center clone name: 200_1_19

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 154794 bases at least Q40

Consensus quality: 160262 bases at least Q30

Consensus quality: 162545 bases at least Q20

Insert size: 171000; agarose-fp

Insert size: 164295; sum-of-contigs

Quality coverage: 4.7 in Q20 bases; agarose-fp

Quality coverage: 4.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 27 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1021: contig of 1021 bp in length
 * 1022 1121: gap of 100 bp
 * 1122 2188: contig of 1067 bp in length
 * 2189 2288: gap of 100 bp
 * 2289 3869: contig of 1581 bp in length
 * 3870 3969: gap of 100 bp
 * 3970 5020: contig of 1051 bp in length
 * 5021 5120: gap of 100 bp
 * 5121 6753: contig of 1633 bp in length
 * 6754 8166: contig of 1313 bp in length
 * 8167 8266: gap of 100 bp
 * 8267 9661: contig of 1395 bp in length
 * 9662 9761: gap of 100 bp
 * 9762 12386: contig of 2625 bp in length
 * 12387 12486: gap of 100 bp
 * 12487 15618: contig of 3132 bp in length
 * 15619 15718: gap of 100 bp

* 15719 18825: contig of 3107 bp in length
 * 18826 18925: gap of 100 bp
 * 18926 22110: contig of 3185 bp in length
 * 22111 22210: gap of 100 bp
 * 22211 25885: contig of 3675 bp in length
 * 25886 25986: gap of 100 bp
 * 25986 31121: contig of 5136 bp in length
 * 31122 31221: gap of 100 bp
 * 31222 35458: contig of 4237 bp in length
 * 35459 35558: gap of 100 bp
 * 35559 40046: contig of 4488 bp in length
 * 40047 40146: gap of 100 bp
 * 40147 47343: contig of 7197 bp in length
 * 47344 47443: gap of 100 bp
 * 47444 52032: contig of 4589 bp in length
 * 52033 52132: gap of 100 bp
 * 52133 57816: contig of 5684 bp in length
 * 57817 57916: gap of 100 bp
 * 57917 63102: contig of 5086 bp in length
 * 63103 63102: gap of 100 bp
 * 63103 70861: contig of 7759 bp in length
 * 70862 70861: gap of 100 bp
 * 70862 76669: contig of 5708 bp in length
 * 76670 76769: gap of 100 bp
 * 76770 84103: contig of 7334 bp in length
 * 84104 84203: gap of 100 bp
 * 84204 91221: contig of 7018 bp in length
 * 91222 91321: gap of 100 bp
 * 91322 100954: contig of 9633 bp in length
 * 100955 101054: gap of 100 bp
 * 101055 109879: contig of 8825 bp in length
 * 109880 109979: gap of 100 bp
 * 109980 126962: contig of 16983 bp in length
 * 126963 127062: gap of 100 bp
 * 127063 166895: contig of 39833 bp in length.

Location/Qualifiers

1. 166895
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3"
 /clone.lib="RP11-200119"
 /clone.lib="RP11-200119"
 1. 1021
 /note="assembly_fragment"
 1122. 12188
 /note="assembly_fragment"
 2289. 3869
 /note="assembly_fragment"
 3970. 5020
 /note="assembly_fragment"
 5121. 6753
 /note="assembly_fragment"
 6854. 8166
 /note="assembly_fragment"
 8267. 9661
 /note="assembly_fragment"
 9762. 12386
 /note="assembly_fragment"
 12487. 15618
 /note="assembly_fragment"
 15719. 18925
 /note="assembly_fragment"
 18926. 22110
 /note="assembly_fragment"
 22211. 25885
 /note="assembly_fragment"
 25986. 31121
 /note="assembly_fragment"
 31222. 35458
 /note="assembly_fragment"
 35559. 40046
 /note="assembly_fragment"

FEATURES

source

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature /note="assembly_fragment"
40147..47343
misc_feature /note="assembly_fragment"
47444..52032
misc_feature /note="assembly_fragment"
vector_end:T7
vector_side:left"
52133..57816
misc_feature /note="assembly_fragment"
57917..63002
misc_feature /note="assembly_fragment"
63103..70861
misc_feature /note="assembly_fragment"
70962..76669
misc_feature /note="assembly_fragment"
76770..84103
misc_feature /note="assembly_fragment"
84204..91221
misc_feature /note="assembly_fragment"
91322..100954
misc_feature /note="assembly_fragment"
101055..109879
misc_feature /note="assembly_fragment"
vector_end:SP6
vector_side:right"
109980..126962
misc_feature /note="assembly_fragment"
127063..166895
misc_feature /note="assembly_fragment"

ORIGIN
Query Match 8.0%; Score 201.6; DB 2; Length 166895;
Best Local Similarity 85.3%; Pred. No. 5e-33;
Matches 249; Conservative 0; Mismatches 39; Indels 4; Gaps 2;
QY 2207 ATATAAAATACCCCGCCAGGCGTGTGAGACCAAGCTTGAATCCAGCATTGGGAG 2266
DB 105201 AATAGCAATGAAGCTGGGCGGTGTGAGACCAAGCTTGAATCCAGCATTGGGAG 105260
QY 2267 GTGGTGGATCACCTGAGGTGGGAGTTCGAGACCAAGCTTGAATCCAGCATTGGGAG 2326
DB 105261 GCAGGTGGATCACCTGAGGTGGGAGTTCGAGACCAAGCTTGAATCCAGCATTGGGAG 105320
QY 2327 CATCTCTACTAAAAACACAAAAATAGCGGCGGTGGGAGTTCGAGACCAAGCTTGAATCCAGCATTGGGAG 2386
DB 105321 CGTCTCTACTAAAAATACAAAAATAGTGGGCGGTGGGAGTTCGAGACCAAGCTTGAATCCAGCATTGGGAG 105380
QY 2387 CTACTCAGAGGCTGAGGC-GGAGAAATTCGTTGAACCCGAGGCTGAGGTTGGTGGCT 2445
DB 105381 CTACTCAGAGGCTGAGGCAGGAAATTCGTTGAACCCGAGGCTGAGGTTGGTGGCT 105437
QY 2446 GAGCTGAGATTGACATTTGCACTCCAGCTGGGCAACAGGAGTAAACCTCC 2497
DB 105438 TACCTGAGATTGGCCATTGTACTCCAGCTGGGCAACAGGAGTAAACCTCC 105489

RESULT 11
AC112202
LOCUS AC112202 170961 bp DNA linear HTG 20-FEB-2002
DEFINITION Homo sapiens chromosome 16 clone RP11-6E10, WORKING DRAFT SEQUENCE,
9 unordered pieces.
AC112202
AC112202.1 GI:18767474
VERSION HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170961)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170961)

AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 408775
Center clone name: RPCI-11_6E10

Summary Statistics
Consensus quality: 163237 bases at least Q40
Consensus quality: 165500 bases at least Q30
Consensus quality: 166457 bases at least Q20
Estimated insert size: 179000; agarose-fp estimation
Estimated insert size: 170161; sum-of-contigs estimation
Quality coverage: 14.19 in Q20 bases; agarose-fp estimation
Quality coverage: 14.93 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1074: contig of 1074 bp in length
* 1075 1174: gap of unknown length
* 1175 2298: contig of 1124 bp in length
* 2299 2398: gap of unknown length
* 2399 3709: contig of 1311 bp in length
* 3710 3809: gap of unknown length
* 3810 5897: contig of 1888 bp in length
* 5898 5797: gap of unknown length
* 5798 7063: contig of 1265 bp in length
* 7063 7162: gap of unknown length
* 7163 10130: contig of 2968 bp in length
* 10131 37572: contig of 27342 bp in length
* 37573 37672: gap of unknown length
* 37673 78655: contig of 40983 bp in length
* 78656 78755: gap of unknown length
* 78756 170961: contig of 92206 bp in length.
FEATURES
Location/Qualifiers
1..170961
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-6E10"
/clone_lib="RPCI human BAC library 11"
ORIGIN
Query Match 8.0%; Score 201.6; DB 2; Length 170961;
Best Local Similarity 85.3%; Pred. No. 5e-33;
Matches 249; Conservative 0; Mismatches 39; Indels 4; Gaps 2;
QY 2207 ATATAAAATACCCCGCCAGGCGTGTGAGACCAAGCTTGAATCCAGCATTGGGAG 2266
DB 93321 AATAGCAATGAAGCTGGGCGGTGTGAGACCAAGCTTGAATCCAGCATTGGGAG 93380
QY 2267 GTGGTGGATCACCTGAGGTGGGAGTTCGAGACCAAGCTTGAATCCAGCATTGGGAG 2326
DB 93381 GCAGGTGGATCACCTGAGGTGGGAGTTCGAGACCAAGCTTGAATCCAGCATTGGGAG 93440
QY 2327 CATCTCTACTAAAAACACAAAAATAGCGGCGGTGGGAGTTCGAGACCAAGCTTGAATCCAGCATTGGGAG 2386
DB 93441 CGTCTCTACTAAAAATACAAAAATAGTGGGCGGTGGGAGTTCGAGACCAAGCTTGAATCCAGCATTGGGAG 93500
QY 2387 CTACTCAGAGGCTGAGGC-GGAGAAATTCGTTGAACCCGAGGAGTGGGAGTGGTGGCT 2445

```
Db 93501 CTACTCAGAGGCTGAGGACGAGAAATTCCTTGAACCCGGAGGTGGAG--GCTGCAGT 93557
QY 2446 GAGCTGAGATTCACCTATTGCACTCCAGCTCGGCAACAGGAGTAAACTCC 2497
Db 93558 TACCTGAGATTGGCCATTTACTCCAGCTGGGCAACAGAGCGAAACTCC 93609

RESULT 12
AC131148/c
LOCUS AC131148.2 174054 bp DNA linear HTG 21-SEP-2002
DEFINITION Homo sapiens clone RP11-622J3, *** SEQUENCING IN PROGRESS ***, 6
unordered pieces.
ACCESSION AC131148
VERSION AC131148.2 GI:23264913
KEYWORDS HTG; HTGS, PHASE1.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174054)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayelle,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,D., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,N.,
Moser,M., Neal,D., Newtonson,J., Newtonson,N., Nguyen,A., Nguyen,N.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,B., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 174054)
Worley,K.C.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (17-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 174054)
Worley,K.C.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department
```

```
COMMENT
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:22296651.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HEKP
Center clone name: RP11-622J3
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 170243 bases at least Q40
Consensus quality: 170621 bases at least Q30
Consensus quality: 170908 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 6 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 4685: contig of 4685 bp in length
* 4686 4785: gap of unknown length
* 4786 9848: contig of 5063 bp in length
* 9849 9948: gap of unknown length
* 9949 30170: contig of 20222 bp in length
* 30171 30270: gap of unknown length
* 30271 51230: contig of 20960 bp in length
* 51231 51330: gap of unknown length
* 51331 76881: contig of 25551 bp in length
* 76882 76981: gap of unknown length
* 76982 174054: contig of 97073 bp in length.
FEATURES
Location/Qualifiers
1..174054
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-622J3"
ORIGIN
Query Match 8.0%; Score 201.6; DB 2; Length 174054;
Best Local Similarity 85.3%; Pred No. 5e-33;
Matches 249; Conservative 0; Mismatches 39; Indels 4; Gaps 2;
QY 2207 ATATAAAATACCCGGCCAGGACGGTGGCTCAGCGCTGTAATCCAGCACTTTGGGAG 2266
Db 14631 AAATAGCAATGTAAGGCTGGGCACGGTGGCTCACCCTGTAATCTGGCACTTTGGAAG 14572
QY 2267 GTGGGTGATCACCCTGAGGTGGGAGTTTCGAGACCAAGCTGACCAAGCATGGTGGAAACCC 2326
Db 14571 CGAGGTGATCACCCTGAGGTGAGGAGTTTCAAGACCAAGCTGGCCCAACATGTTAAACCC 14512
QY 2327 CATCTCTACTAAACACAAAAAATTAGCCGGCGCTGGTGGCACATGCTGTAAATCCCAAG 2386
Db 14511 CTTCTCTACTAAATTAACAAAAATAGCTGGCGTGGTGGGACATGCTGTAAATCCCAAG 14452
QY 2387 TTACTCAGGAGCTGAGGC-GGAGAAATGCTTTGAACCCGGAGGTGGAGGTTTGGCGGT 2445
Db 14451 CTACTCAGGAGCTGAGGCAGGAGAAATGCTTGAACCCGGAGGTGGAG--GCTGCAGT 14395
QY 2446 GAGCTGAGATTGCATATTGCACTCCAGCTGGGCAACAGGAGTAAACTCC 2497
Db 14394 TACCTGAGATTGGCGCATTTGTACTCCAGCCTGGGCAACAGAGCGAAACTCC 14343
```

```

RESULT 13
AF168787
LOCUS
DEFINITION
Homo sapiens vanilloid receptor gene, partial sequence; CARKL and
CTNS genes, complete cds; TIP1 gene, partial cds; P2X5b and P2X5a
genes, complete cds; and HUMINAE gene, partial cds.
ACCESSION
AF168787
VERSION
AF168787.1 GI:7239175
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 20237)
Touchman,J.W., Ankster,Y., Dietrich,N.L., Maduro,V.V.,
McDowell,G., Shotelersuk,V., Bouffard,G.G.,
Beckstrom-Sternberg,S.M., Gahl,W.A. and Green,E.D.
The genomic region encompassing the nephropathic cystinosis gene
(CTNS): complete sequencing of a 200-kb segment and discovery of a
novel gene within the common cystinosis-causing deletion
Genome Res. 10 (2), 165-173 (2000)
2 (bases 1 to 20237)
Dietrich,N.L., Ferguson,M., Sorbello,E., Gupta,A., Torkzadeh,R.,
Varner,C., Walker,M., Bouffard,G.G., Beckstrom-Sternberg,S.M. and
Touchman,J.W.
Direct Submission
Submitted (14-JUL-1999) Intramural Sequencing Center, National
Institutes of Health, 8717 Grovemont Circle, Gaithersburg, Maryland
20877, USA
FEATURES
source
1. .200237
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p13"
complement(<1. .27738)
/genes="vanilloid receptor"
/notes="coding sequence not determined; similar to Rattus
norvegicus vanilloid receptor subtype 1 encoded by GenBank
Accession Number AF029310"
evidence=not_experimental
5. .79
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
129. .314
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
1252. .1851
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
2437. .2736
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
3093. .3142
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
3389. .3685
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
4117. .4246
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
4254. .4625
repeat_region
/rpt_family="L1"
/rpt_type=dispersed
4616. .4784
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
5466. .5735
repeat_region
/rpt_family="Alu"
/rpt_type=dispersed
complement(6071. .6363)
/rpt_family="Alu"
7935. .8018
/rpt_family="MIR"
/rpt_type=dispersed
8055. .8343
/rpt_family="Alu"
/rpt_type=dispersed
10007. .10275
/rpt_family="Alu"
/rpt_type=dispersed
10255. .10507
/rpt_family="MER42"
/rpt_type=dispersed
10549. .10838
/rpt_family="Alu"
/rpt_type=dispersed
10834. .11217
/rpt_family="MER42"
/rpt_type=dispersed
11289. .12215
/rpt_family="Alu"
/rpt_type=dispersed
13386. .13659
/rpt_family="Alu"
/rpt_type=dispersed
13735. .13906
/rpt_family="Alu"
/rpt_type=dispersed
14075. .14295
/rpt_family="L1"
/rpt_type=dispersed
complement(14848. .15120)
/rpt_family="Alu"
15173. .15439
/rpt_family="MER20"
/rpt_type=dispersed
16062. .16377
/rpt_family="Alu"
/rpt_type=dispersed
complement(16495. .16633)
/rpt_family="Alu"
16960. .17705
/rpt_family="Alu"
/rpt_type=dispersed
17979. .18328
/rpt_family="Alu"
/rpt_type=dispersed
19355. .19786
/rpt_family="Alu"
/rpt_type=dispersed
19869. .20162
/rpt_family="Alu"
/rpt_type=dispersed
complement(20503. .20804)
/rpt_family="Alu"
21844. .22129
/rpt_family="Alu"
/rpt_type=dispersed
22918. .23184
/rpt_family="Alu"
/rpt_type=dispersed
23880. .24184
/rpt_family="Alu"
/rpt_type=dispersed
24368. .24421
/rpt_family="Alu"

```

repeat_region	/rpt_type=dispersed complement(24564..24833) /rpt_family="Alu" /rpt_type=dispersed complement(28469..28625) /rpt_family="Alu" /rpt_type=dispersed 29167..29464 /rpt_family="Alu" /rpt_type=dispersed complement(30577..30675) /rpt_family="MIR" /rpt_type=dispersed complement(32993..33290) /rpt_family="Alu" /rpt_type=dispersed complement(33885..34180) /rpt_family="Alu" /rpt_type=dispersed 34276..34562 /rpt_family="Alu" /rpt_type=dispersed complement(34947..35225) /rpt_family="Alu" /rpt_type=dispersed 35626..35919 /rpt_family="Alu" /rpt_type=dispersed 36253 /note="cystinosis deletion breakpoint" /evidence=experimental complement(36925..37208) /rpt_family="Alu" /rpt_type=dispersed 38170..38470 /rpt_family="Alu" /rpt_type=dispersed 39628..39934 /rpt_family="Alu" /rpt_type=dispersed 40422..40545 /rpt_type="MIR" /rpt_family="MIR" 40570..40839 /rpt_family="Alu" /rpt_type=dispersed complement(41304..41558) /rpt_family="Alu" 42368..42675 /rpt_type=dispersed complement(42857..43141) /rpt_family="Alu" /rpt_type=dispersed 43262..43546 /rpt_family="Alu"	QY	2381	TCCAGCTACTCAGGAGGCTGAGGC-GGAGAAATTGCTTGAACCCGGAAGGTGGAGGTTGT	2439
repeat_region		Db	137840	TCCAGCTACTCGGGAGGCTGAGGCAGGAGAAATTGCTTGAACCCAGGAGGACAGGTTGT	137899
repeat_region		QY	2440	TGCGGTGAGCTGAGATTGCACATTATTCACCTCCAGCCTGGGCAACAGGAGTAAACTCCCC	2499
repeat_region		Db	137900	---GGTGAGCTGAGATGACGCCATTGCACCTCCAGCCTGGGCAACAGAGCAAAACTCTGC	137956
repeat_region		QY	2500	CCCA 2503	
repeat_region		Db	137957	CTCA 137960	
repeat_region		RESULT 14			
repeat_region		AC009301/c			
repeat_region		LOCUS			
repeat_region		DEFINITION Homo sapiens BAC clone RP11-62F14 from 2, complete sequence.			
repeat_region		AC009301			
repeat_region		AC009301.3 GI:7243885			
repeat_region		HTG.			
repeat_region		SOURCE			
repeat_region		Homo sapiens (human)			
repeat_region		ORGANISM			
repeat_region		Homo sapiens			
repeat_region		Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
repeat_region		1 (bases 1 to 129625)			
repeat_region		Sulston,J.E. and Waterston,R.			
repeat_region		Toward a complete human genome sequence			
repeat_region		Genome Res. 8 (11), 1097-1108 (1998)			
repeat_region		99063792			
repeat_region		PUBMED			
repeat_region		9847074			
repeat_region		2 (bases 1 to 129625)			
repeat_region		Drone,K., Leonard,S. and Doebber,A.			
repeat_region		The sequence of Homo sapiens BAC clone RP11-62F14			
repeat_region		Unpublished			
repeat_region		REFERENCE			
repeat_region		3 (bases 1 to 129625)			
repeat_region		Waterston,R.H.			
repeat_region		Direct Submission			
repeat_region		Submitted (13-AUG-1999)			
repeat_region		Genome Sequencing Center, Washington			
repeat_region		University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
repeat_region		MO 63108, USA			
repeat_region		REFERENCE			
repeat_region		4 (bases 1 to 129625)			
repeat_region		Waterston,R.H.			
repeat_region		Direct Submission			
repeat_region		Submitted (15-MAR-2000)			
repeat_region		Genome Sequencing Center, Washington			
repeat_region		University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
repeat_region		MO 63108, USA			
repeat_region		REFERENCE			
repeat_region		5 (bases 1 to 129625)			
repeat_region		Waterston,R.			
repeat_region		Direct Submission			
repeat_region		Submitted (30-SEP-2000)			
repeat_region		Department of Genetics, Washington			
repeat_region		University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA			
repeat_region		On Mar 15, 2000 this sequence version replaced gi:6598926.			
repeat_region		----- Genome Center			
repeat_region		Center: Washington University Genome Sequencing Center			
repeat_region		Center code: WUGSC			
repeat_region		Web site: http://genome.wustl.edu/gsc			

Query Match 7.9%; Score 198.8; DB 9; Length 200237;
Best Local Similarity 87.5%; Pred.No.1.9e-32;
Matches 266; Conservative 0; Mismatches 27; Indels 11; Gaps 4;

QY 2207 ATATATAAATACCCGGCCAGGACGGTGGCTCAGCCCTGTATCCAGCACTTTGG--- 2263
Db 137661 ACATTAATAAACCAGCCGGCATGTGTGCTCAGCCCTGTATCCAGCACTTTGGGAG 137720

QY 2264 ---GAGTGGTGGATCAGCTGAGGTGGGAGTTGAGACCGCTGACAGCATGGTGG 2320
Db 137721 GCCGAGGTGGTGGATCAGCTGAGGTTCGAGACCGCTGACCAACATGG-AG 137779

QY 2321 AACCCCCATCTCTACTATAAACAACAAAAATTAGCCGGCGCTGGTGGCACATGCCCTGTAA 2380
Db 137780 AAACCCCATCTCTACTATAAATAACAAAAAGTAGCCGGAGCGTGGTGGCACATGCCCTGTAA 137839

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, B., Tateno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.regen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-153B21, 200 bp overlap. Actual start of this clone is at base position 133171 of RP11-153B21; actual end is at base position 129625 of RP11-62F14.

FEATURES

source	1. .129625	Location/Qualifiers
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="2"	
	/map="2"	
	/clone="RP11-62F14"	
	/clone_lib="RPCI-11"	
misc_feature	1539..1647	
	/notes="match to EST AA358669 (NID:g2010987)"	
repeat_region	1752..1921	
	/rpt_family="MIR"	
repeat_region	2556..2676	
	/rpt_family="MIR"	
repeat_region	3904..4038	
	/rpt_family="MIR"	
repeat_region	4064..4161	
	/rpt_family="MIR"	
repeat_region	5485..5701	
	/rpt_family="MIR"	
repeat_region	5766..5858	
	/rpt_family="MER1_type"	
repeat_region	5859..6170	
	/rpt_family="Alu"	
repeat_region	6171..6270	
	/rpt_family="MER1_type"	
repeat_region	6302..6394	
	/rpt_family="L2"	
repeat_region	6369..6482	
	/rpt_family="MIR"	
misc_feature	7024..7525	
	/notes="match to EST AAL179680 (NID:g1761913) zpl3e05.r1"	
misc_feature	7845..7846	
	/notes="match to EST AAL176625 (NID:g1757748) zpl3e05.s1"	
misc_feature	7931..8410	
	/notes="match to EST AAL176625 (NID:g1757748) zpl3e05.s1"	
repeat_region	8170..8362	
	/rpt_family="MIR"	
repeat_region	8405..8452	
	/rpt_family="L2"	
repeat_region	8428..8707	
	/rpt_family="L2"	
repeat_region	8925..9219	
	/rpt_family="L2"	
misc_feature	9387..9454	
	/notes="match to EST AA358669 (NID:g2010987)"	
repeat_region	10445..10657	
	/rpt_family="MIR"	
repeat_region	11048..11344	
	/rpt_family="Alu"	
repeat_region	11348..11496	
	/rpt_family="L1"	
repeat_region	11498..11520	
	/rpt_family="AT_rich"	
misc_feature	11874..11941	
	/notes="match to EST AA358669 (NID:g2010987)"	
misc_feature	12150..12224	
	/notes="match to EST AA285241 (NID:g1929551)"	
repeat_region	12358..12434	
	/rpt_family="MIR"	
repeat_region	12805..12833	
	/rpt_family="AT_rich"	
repeat_region	12939..13014	
	/rpt_family="MIR"	
repeat_region	13382..13500	
	/rpt_family="Alu"	
repeat_region	13603..13939	
	/rpt_family="L2"	
repeat_region	14032..14081	
	/rpt_family="L2"	
repeat_region	14244..14554	
	/rpt_family="Alu"	
repeat_region	15087..15575	
	/rpt_family="L1"	
repeat_region	16160..16221	
	/rpt_family="(TG)n"	
repeat_region	16224..16856	
	/rpt_family="L1"	
repeat_region	17784..17926	
	/rpt_family="MIR"	
repeat_region	19043..19564	
	/rpt_family="MER21-group"	
repeat_region	20471..21063	
	/rpt_family="L2"	
repeat_region	21165..21598	
	/rpt_family="MaLR"	
misc_feature	21972..22068	
	/notes="match to EST AA285241 (NID:g1929551)"	
misc_feature	22909..22993	
	/notes="match to EST AA285241 (NID:g1929551)"	
misc_feature	22935..22993	
	/notes="match to EST AA081997 (NID:g1624063) zn26d08.r1"	
repeat_region	24647..25035	
	/rpt_family="MaLR"	
misc_feature	25547..25650	
	/notes="match to EST AA081997 (NID:g1624063) zn26d08.r1"	
misc_feature	25547..25576	
	/notes="match to EST AA285241 (NID:g1929551)"	
repeat_region	26490..26608	
	/rpt_family="L2"	
repeat_region	29443..29466	
	/rpt_family="(GA)n"	
repeat_region	29466..29491	
	/rpt_family="(TG)n"	
repeat_region	29522..29561	
	/rpt_family="(TG)n"	
repeat_region	29899..29989	
	/rpt_family="MIR"	
repeat_region	30000..30490	
	/rpt_family="MaLR"	
repeat_region	31975..34081	
	/rpt_family="MIR"	
repeat_region	34579..34750	
	/rpt_family="MER1_type?"	
STS	35327..35502	
	/db_xref="GI:4529205"	
repeat_region	35460..35512	
	/rpt_family="MIR"	
misc_feature	35657..35783	
	/notes="match to EST AA081997 (NID:g1624063) zn26d08.r1"	

Query Match

Best Local Similarity 7.9%; Score 198.6; DB 9; Length 129625; Pred. No. 2.3e-32;

Matches 252; Conservative 0; Mismatches 44; Indels 7; Gaps 2;

QY 2207 ATATATAAATACCCCGCCAGCAGCGTGGCTCACGCCCTGTATCCAGCAGCTTTGGG-- 2264

Db 104230 ATTATGCTCAAGGCCGGGAGCAGTACCCATGCTGTATCCAGCAGCTTTGGGAG 104171

QY 2265 ----AGTGGGTGATCACTGAGTGGGAGTTCGAGACCGCCTGACGAGCATGGTGG 2320

Db 104170 GCCAAGTGGCAGATCACTGAGTTCGAGAGTTCGAGACCGCCTGGCTAAATGGTGA 104111

QY 2321 AACCCCATCTCTACTTAAAAACACAAAAATTAGCCGGCGGTGGTGCACATGCCCTGTAA 2380

Db 104110 AAC-CCATTTCTACTTAAAAATACAAAAATTAGCCAGCGGTGGTGCACAGCCTGTAA 104052

QY 2381 TCCAGCTACTCAGGAGGCTGAGCGCGGAGAAATGCTTGAACCGGAGGTGGAGGTGTGT 2440

Db 104051 TCTCAGCTACTCAGGAGGCTGAGGAGGAGATCGCTTGAACCTGGAGGTGGAGTTGCA 103992

QY 2441 GCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAACTCCCC 2500

Db 103991 GCAGTGAGCTGAGATTGTGCTATTGCACTCCAGCCTGGGCAACAGGAGAACTCCATC 103932

QY 2501 CCA 2503

Db 103931 TCA 103929

RESULT 15

AC107422

LOCUS AC107422

DEFINITION Homo sapiens chromosome 3 clone RP11-372J10, WORKING DRAFT

SEQUENCE, 14 unordered pieces.

AC107422

VERSION AC107422.3 GI:20335954

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 318586)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-oman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayete,M., Banks,T., Barbarta,J., Benton,J., Biemage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowls,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,W.D., Dathorne,S.R., David,R., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flegg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Hosi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,B., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S., Ogum,K., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Roife,W., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,

Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umami,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 318586)

Worley,K.C.

Direct Submission

Submitted (20-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 318586)

Worley,K.C.

Direct Submission

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Apr 28, 2002 this sequence version replaced gi:18449820.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HDRW

Center clone name: RP11-372J10

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 306871 bases at least Q40

Consensus quality: 310966 bases at least Q30

Consensus quality: 313294 bases at least Q20

Estimated insert size: 170415; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 14 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 4906: contig of 4906 bp in length

4907 5006: gap of unknown length

5007 12370: contig of 7364 bp in length

12371 12470: gap of unknown length

12471 21459: contig of 8989 bp in length

21460 21559: gap of unknown length

21560 33615: contig of 12056 bp in length

33616 33715: gap of unknown length

33716 47492: contig of 13777 bp in length

47493 69615: contig of 22023 bp in length

69616 69715: gap of unknown length

69716 95747: contig of 26032 bp in length

95748 121218: contig of 25371 bp in length

121219 121318: gap of unknown length

121319 146590: contig of 25272 bp in length

146591 146690: gap of unknown length

146691 168128: contig of 21438 bp in length

168129 168228: gap of unknown length

168229 194928: contig of 28699 bp in length

194928 195027: gap of unknown length

195028 221744: contig of 26717 bp in length

* 221745 221844: gap of unknown length
 * 221845 267827: contig of 45983 bp in length
 * 267828 267927: gap of unknown length
 * 267928 318586: contig of 50659 bp in length.

FEATURES

Location/Qualifiers
 source 1..318586
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-372J10"

ORIGIN

Query Match 7.98; Score 198.6; DB 2; Length 318586;
 Best Local Similarity 83.2%; Pred. No. 2e-32;
 Matches 252; Conservative 0; Mismatches 44; Indels 7; Gaps 2;
 QY 2207 ATATAAAATACCCGCGCAGGCACGCTGCTCACGCTGTATCCAGCAGCTTTGGG-- 2264
 |||||
 Db 242665 ATTTAATGGCTCAAGGCGGGAGCAGTGACCCATGCCTGTATCCAGCAGCTTTGGGAG 242724
 |||||
 QY 2265 ----AGGTGGTGGATCACCTGAGGTGGGAGTTCCAGACCCGCTGACCCAGCATGGTGG 2320
 |||||
 Db 242725 GCCAAGGTGGGACAGATCACCTGAGGTGAGGATTCGAGACCCAGCTGGCTAACATGGTGA 242784
 |||||
 QY 2321 AACCCCATCTCTACTATAAAACACAAATTTAGCCGGCGTGGTGGCACATGCCCTGTAA 2380
 |||||
 Db 242785 AAC-CCCATTTCTACTAAAAAATACAAAAAATTTAGCCAGGCGTGGTGGCACACGCCCTGTAA 242843
 |||||
 QY 2381 TCCAGCTACTCAGGAGGCTGAGGCGGAGAAATTTGTTGAACCCGGAAGGTGGAGGTTGTT 2440
 |||||
 Db 242844 TCTCAGCTACTCAGGAGGCTGAGGAGGAGATCGCTTGAACCTGGGAGGTGGAGGTTGCA 242903
 |||||
 QY 2441 GCGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCC 2500
 |||||
 Db 242904 GCAGTGAGCTGAGATTGTGCAATTGCATTGCACCTCCAGCCTGGGCAACAGAGAGAAACTCCATC 242963
 |||||
 QY 2501 CCA 2503
 |||||
 Db 242964 TCA 242966

Search completed: September 9, 2005, 03:30:41
 Job time : 7229.71 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 8, 2005, 22:23:54 ; Search time 886.428 Seconds
(without alignments)
16735.564 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gacctaataatatagaggt.....agtaaaactccccccacc 2506

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2505.6	100.0	2506	3 AAZ87584	Aaz87584 Prostate
2	2494	99.5	2505	4 AAS04001	Aas04001 Biomarker
3	1961.6	78.3	2088	3 AAZ87583	Aaz87583 Prostate
4	1950	77.8	2087	4 AAS04000	Aas04000 Biomarker
5	626.2	25.0	757	2 AAV16883	Aav16883 Human pro
6	626.2	25.0	757	3 AAX26018	Aax26018 Prostate
7	626.2	25.0	757	3 AAZ87503	Aaz87503 Prostate
8	626.2	25.0	757	4 AAS03722	Aas03722 Biomarker
9	190.6	7.6	299598	12 ADQ59380	Adq59380 Human can
10	186.8	7.5	215974	12 ADQ97523	Adq97523 Human can
11	186.4	7.4	2680	4 AAK85292	Aak85292 Human inn
12	186.4	7.4	5788	5 ABA15481	Abal5481 Human ner
13	186.4	7.4	5788	5 ABA15482	Abal5482 Human ner
14	186.4	7.4	14417	4 AAK87584	Aak87584 Human inn
15	186.4	7.4	14417	4 AAK71816	Aak71816 Human inn
16	186.4	7.4	14417	4 AAK73113	Aak73113 Human inn
17	186.4	7.4	14417	4 AAI62923	Aai62923 Human gen
18	186.4	7.4	14417	8 ABZ74674	Abz74674 Secretd
19	186.4	7.4	14417	10 ABZ68196	Abz68196 Human sec
20	186.4	7.4	14426	4 AAK73099	Aak73099 Human inn

C	21	186.4	7.4	14426	4	AAK87568	Aak87568 Human inn
	22	186.4	7.4	14426	4	AAK85290	Aak85290 Human inn
	23	186.4	7.4	14426	4	AAK71814	Aak71814 Human inn
	24	186.4	7.4	14426	4	AAI62921	Aai62921 Human gen
C	25	186.4	7.4	14426	8	ABZ74673	Abz74673 Secretd
C	26	186.4	7.4	14426	10	ABZ68195	Abz68195 Human sec
	27	184.4	7.4	874	4	AAH98452	Aah98452 Human EST
	28	184.4	7.4	1599	10	ADG10849	Adg10849 Human STA
C	29	183.8	7.3	41150	10	ADL13819	Adl13819 Osteoarth
	30	183.8	7.3	44348	12	ADN48556	Adn48556 Human Not
	31	183.6	7.3	106746	3	AAAI0225	Aaa10225 Human PCT
C	32	183.6	7.3	108316	10	ADC87336	Adc87336 Human GPC
C	33	183.4	7.3	5286	13	ADS89008	Ads89008 Human GGT
C	34	183.2	7.3	8846	4	ABA07221	Abas07221 Human pan
C	35	183.2	7.3	8846	4	AAK89858	Aak89858 Human dig
	36	183.2	7.3	110665	12	ADQ19183	Adq19183 Human sof
	37	183.2	7.3	227968	6	ABK83497	Abk83497 Human CDN
	38	183.2	7.3	227968	12	ADQ18538	Adq18538 Human sof
	39	183	7.3	2952	6	ABK69090	Abk69090 DNA encod
C	40	183	7.3	73723	13	ABD33145	Abd33145 Human can
	41	183	7.3	270150	11	ADP65796	Adp65796 Human 16p
	42	182.8	7.3	463	9	ACH36634	Ach36634 Human end
	43	182.6	7.3	89625	11	ACN45194	Acn45194 Human gen
C	44	182.2	7.3	71048	11	ACN44082	Acn44082 Human gen
C	45	181.8	7.3	91760	11	ACN44410	Acn44410 Human gen

ALIGNMENTS

RESULT 1
AAZ87584
ID AAZ87584 standard; DNA; 2506 BP.
XX
AC AAZ87584;
DT 19-APR-2000. (first entry)
XX
DE Prostate disease marker UC Band #28 splice variant.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN WO9964631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1999; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
XX
PT P-PSDB; AAY59296.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Claim 1; Page 184-186; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer.
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid

CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences from various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX

SQ Sequence 2506 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 1 Other;

Query Match		100.0%;	Score 2505.6;	DB 3;	Length 2506;		
Best Local Similarity		100.0%;	Pred. No. 0;				
Matches 2506;		Conservative	0;	Mismatches	0;	Indels	0; Gaps
QY	1	GACCTTAATATATATCGAGTGGCTGAATGATGATATTAATTTACAAAATTAATCTTTCTA	60				
Db	1	GACCTTAATATATATCGAGTGGCTGAATGATGATATTAATTTACAAAATTAATCTTTCTA	60				
QY	61	TTGGCTACAGAGCTACAATTCATTTACAGTAGGCCACCATGAGGCTCTTTAAGGAACC	120				
Db	61	TTGGCTACAGAGCTACAATTCATTTACAGTAGGCCACCATGAGGCTCTTTAAGGAACC	120				
QY	121	AGAAATATGAGGATATGCACAATATTAATTCACATTTTACAGATCAGAAAATTTGAGGCACA	180				
Db	121	AGAAATATGAGGATATGCACAATATTAATTCACATTTTACAGATCAGAAAATTTGAGGCACA	180				
QY	181	GATTAAGTAATCTCCCAAGGCTACCAAGGCATTTAGCTCCAGAAAATCTGTCTTTACCAT	240				
Db	181	GATTAAGTAATCTCCCAAGGCTACCAAGGCATTTAGCTCCAGAAAATCTGTCTTTACCAT	240				
QY	241	TCGTCTACAGGTATTTCCAAAAGAAAGAAAGTAAAGAAAGTCAAAGGCAACAGAGT	300				
Db	241	TCGTCTACAGGTATTTCCAAAAGAAAGAAAGTAAAGAAAGTCAAAGGCAACAGAGT	300				
QY	301	TCATTTGATTAATTCATAGAAACAGTACACCATGCAATTCACACCTTTGCAGACACACT	360				
Db	301	TCATTTGATTAATTCATAGAAACAGTACACCATGCAATTCACACCTTTGCAGACACACT	360				
QY	361	TGACCATGAAGGTTCTCCCAATGAATGTTCTCTCAATTAATCTTTCAGAAGCCATATTATCA	420				
Db	361	TGACCATGAAGGTTCTCCCAATGAATGTTCTCTCAATTAATCTTTCAGAAGCCATATTATCA	420				
QY	421	CATTGACCTTTCAGTTAACTCAGACCCCTAGGCTGGAATGCTCTCTCTACTTATCCA	480				
Db	421	CATTGACCTTTCAGTTAACTCAGACCCCTAGGCTGGAATGCTCTCTCTACTTATCCA	480				
QY	481	AAACTATACATCCACAGATCATATAAATCTCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA	540				
Db	481	AAACTATACATCCACAGATCATATAAATCTCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA	540				
QY	541	ATAAAAATGGTTGAAAGGCAATCTGCTACCAATGACTGTTTAAAGCCCAAGTAAC	600				
Db	541	ATAAAAATGGTTGAAAGGCAATCTGCTACCAATGACTGTTTAAAGCCCAAGTAAC	600				
QY	601	TGAACCATTTCCAACTTCAATTTTACATTATGAAAAGAAATTTGATGATGTAGGAGTTATTTC	660				
Db	601	TGAACCATTTCCAACTTCAATTTTACATTATGAAAAGAAATTTGATGATGTAGGAGTTATTTC	660				
QY	661	AAATCTTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACATCATAGATTTATTCT	720				
Db	661	AAATCTTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACATCATAGATTTATTCT	720				
QY	721	ATTATCTCAATTTAGTTGTTATTTATCCTAGTGGGCCATTTAAAACTACCAATGTGTT	780				
Db	721	ATTATCTCAATTTAGTTGTTATTTATCCTAGTGGGCCATTTAAAACTACCAATGTGTT	780				
QY	781	TCGTGCTCTCCATTAAGTCAATTAACATAACGAGCAATTAAGTAAAGCATGTGCCAGAT	840				
Db	781	TCGTGCTCTCCATTAAGTCAATTAACATAACGAGCAATTAAGTAAAGCATGTGCCAGAT	840				

QY	841	GCTCGCTAGGCACACAGAGGATAAAAAATACTATTAGTATATACCACTAAATTTTCGCTT	900
Db	841	GCTCGCTAGGCACACAGAGGATAAAAAATACTATTAGTATATACCACTAAATTTTCGCTT	900
QY	901	AGTAACTAGTGAAATGTTTCAAGTCAATGCTCAGTCAAGAGTTGAGGACACATTACAATGT	960
Db	901	AGTAACTAGTGAAATGTTTCAAGTCAATGCTCAGTCAAGAGTTGAGGACACATTACAATGT	960
QY	961	GTAATGGAACCAAGGAAGTGAACCTTTGGATAGTGGGACCTAGTGTATTATATATT	1020
Db	961	GTAATGGAACCAAGGAAGTGAACCTTTGGATAGTGGGACCTAGTGTATTATATATT	1020
QY	1021	TAAATGATTTCTGACTCTATCATTTGGCCCTCAAAAACAGAGTTGTCTTTTCTTGGTTTT	1080
Db	1021	TAAATGATTTCTGACTCTATCATTTGGCCCTCAAAAACAGAGTTGTCTTTTCTTGGTTTT	1080
QY	1081	GTTTTCTTCAATATGGAATCTTTGTGCCAGCACAGTGCCTGACACATAGAAAAACAATC	1140
Db	1081	GTTTTCTTCAATATGGAATCTTTGTGCCAGCACAGTGCCTGACACATAGAAAAACAATC	1140
QY	1141	AATATTTCTGAATAAATGATTTAAAAATCAGAGAACTTTCCCATCTCTGTTTGGATCTAT	1200
Db	1141	AATATTTCTGAATAAATGATTTAAAAATCAGAGAACTTTCCCATCTCTGTTTGGATCTAT	1200
QY	1201	AGAAATATGAGGATGATGAGGCTCTGCAATTTATATGCGCTTAAATTAAGATTAT	1260
Db	1201	AGAAATATGAGGATGATGAGGCTCTGCAATTTATATGCGCTTAAATTAAGATTAT	1260
QY	1261	GTGAGAAAGTTTAAAGACACATTTAGTAGAGTGAATTTGAAATATATAGTAAACAATCTTG	1320
Db	1261	GTGAGAAAGTTTAAAGACACATTTAGTAGAGTGAATTTGAAATATATAGTAAACAATCTTG	1320
QY	1321	TGTTGGTCTTTAAAAAGATATTAATAGTAAATGAAATCTCCATCTCAAAAATAATG	1380
Db	1321	TGTTGGTCTTTAAAAAGATATTAATAGTAAATGAAATCTCCATCTCAAAAATAATG	1380
QY	1381	CATAAATCTTTTAAAGGAAATCAATCTCAGGCTTTCAATGTTTGTTCATTTACTTTTT	1440
Db	1381	CATAAATCTTTTAAAGGAAATCAATCTCAGGCTTTCAATGTTTGTTCATTTACTTTTT	1440
QY	1441	CATATATTTTACCATCTGCTGAAGGAGTCAATATCAAAGGGTAAAGAAAGATGGAGGA	1500
Db	1441	CATATATTTTACCATCTGCTGAAGGAGTCAATATCAAAGGGTAAAGAAAGATGGAGGA	1500
QY	1501	AAACTCAGTAAAGATTAATTTAGTCTGTTGCAAGTAGAAAAAGATTCTCATCACTCAA	1560
Db	1501	AAACTCAGTAAAGATTAATTTAGTCTGTTGCAAGTAGAAAAAGATTCTCATCACTCAA	1560
QY	1561	CCTTATGACGAGGAGAGGAGGCTGTTTGAGAACCATTTTACCTTAGCAGAACCAATAT	1620
Db	1561	CCTTATGACGAGGAGAGGAGGCTGTTTGAGAACCATTTTACCTTAGCAGAACCAATAT	1620
QY	1621	TTTAGACACTTCCCTGCAATTAATCTGCAAAAACATATGTTTGCAAAACCTGTTTGTGATCAAC	1680
Db	1621	TTTAGACACTTCCCTGCAATTAATCTGCAAAAACATATGTTTGCAAAACCTGTTTGTGATCAAC	1680
QY	1681	CTCCAAACACGACACATTTGAGGTTAAATTTTTCATCAAAACATTTGATTTTCTTTA	1740
Db	1681	CTCCAAACACGACACATTTGAGGTTAAATTTTTCATCAAAACATTTGATTTTCTTTA	1740
QY	1741	ACGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT	1800
Db	1741	ACGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT	1800
QY	1801	TTCTCCCTGTTTATAGCAGCAAGCAAAATTAGCCATTTTCTCTCAAACTTCACTAAATGA	1860
Db	1801	TTCTCCCTGTTTATAGCAGCAAGCAAAATTAGCCATTTTCTCTCAAACTTCACTAAATGA	1860
QY	1861	TCACATTTCTTCCAAAAGGAACTCTAGAGAACCCAAATGCCGAGTTAAGACATCAAAA	1920
Db	1861	TCACATTTCTTCCAAAAGGAACTCTAGAGAACCCAAATGCCGAGTTAAGACATCAAAA	1920
QY	1921	CTAACCATCTGAAGAAACTTCCCAAGTGAAGACTCTGCCATTAATAAACATTTACCGAGAGG	1980

QY 661 AATTCCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACCTCATAGATTATTATCT 720
DB |||||
QY 661 AATTCCTAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACCTCATAGATTATTATCT 720
DB |||||
QY 721 ATTATCTCAATTTAGTTTGTATTATATCTAGTGGGCCATTTAAACACTACCATGTGTT 780
DB |||||
QY 721 ATTATCTCAATTTAGTTTGTATTATATCTAGTGGGCCATTTAAACACTACCATGTGTT 780
DB |||||
QY 781 TCTGTCTCTCCATTAAGTCAATTAACCTAACTAAACGAGCAATTAAGTAAAGCCATGTCCAGAT 840
DB |||||
QY 781 TCTGTCTCTCCATTAAGTCAATTAACCTAACTAAACGAGCAATTAAGTAAAGCCATGTCCAGAT 840
DB |||||
QY 841 GCTCGCTAGGCACGAGGGGATAAAAACAATACCTTATAGTATACCACTAAATTTTCGCTT 900
DB |||||
QY 841 GCTCGCTAGGCACGAGGGGATAAAAACAATACCTTATAGTATACCACTAAATTTTCGCTT 900
DB |||||
QY 901 AGTAAGTGTGAATTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTAACAATGT 960
DB |||||
QY 901 AGTAAGTGTGAATTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTAACAATGT 960
DB |||||
QY 961 GTAATGGAACCAAGGAGTGAATCTTTGGATTAAGTGGGACTAGTGTATTATATATT 1020
DB |||||
QY 961 GTAATGGAACCAAGGAGTGAATCTTTGGATTAAGTGGGACTAGTGTATTATATATT 1020
DB |||||
QY 1021 TAATTTGATTTCTGACTCTATCAATTTGGCTCCAAAACACAGATTGTGTTTTCTTTGGTTTT 1080
DB |||||
QY 1021 TAATTTGATTTCTGACTCTATCAATTTGGCTCCAAAACACAGATTGTGTTTTCTTTGGTTTT 1080
DB |||||
QY 1081 GTTTTCTTCACATATGGGATCTTCTGTGCCCAGACAGTGCCTGACACATAGAAAACAATC 1140
DB |||||
QY 1081 GTTTTCTTCACATATGGGATCTTCTGTGCCCAGACAGTGCCTGACACATAGAAAACAATC 1140
DB |||||
QY 1141 AATATTTGCTGAATTAATTAATAAATAACAGAACTTTCCCATTTCTGTTTGGATCTAT 1200
DB |||||
QY 1141 AATATTTGCTGAATTAATTAATAAATAACAGAACTTTCCCATTTCTGTTTGGATCTAT 1200
DB |||||
QY 1201 AGAACATCCAGAGTAGTGAATGAGGCTCTGCAATTTATATGCTTTAAATTAAGATTAT 1260
DB |||||
QY 1201 AGAACATCCAGAGTAGTGAATGAGGCTCTGCAATTTATATGCTTTAAATTAAGATTAT 1260
DB |||||
QY 1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGAATTTGAAATATATAGTAAACACTTGGAAA 1320
DB |||||
QY 1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGAATTTGAAATATATAGTAAACACTTGGAAA 1320
DB |||||
QY 1321 TCGTGTGCTTTTAAAGAGATATTAATAGATAATATGAAATCTCCATCTCAAAAATAATG 1380
DB |||||
QY 1321 TCGTGTGCTTTTAAAGAGATATTAATAGATAATATGAAATCTCCATCTCAAAAATAATG 1380
DB |||||
QY 1381 CATAAATCTTTTAAAGGAAAAATCACATCTCCAGGCTTTCAATGTTTCAATCTTTT 1440
DB |||||
QY 1381 CATAAATCTTTTAAAGGAAAAATCACATCTCCAGGCTTTCAATGTTTCAATCTTTT 1440
DB |||||
QY 1441 CATATATTTTAAAGGAAATCTGCTGAAGCAGTCAATATCAAAAGGTAAAGAAAGATGGAGGA 1500
DB |||||
QY 1441 CATATATTTTAAAGGAAATCTGCTGAAGCAGTCAATATCAAAAGGTAAAGAAAGATGGAGGA 1500
DB |||||
QY 1501 ABACTCAGTAAGAAATTTATATTAGTCTGTTTGAAGTGAAGAAAGATTTCTCATCACTCAA 1560
DB |||||
QY 1501 ABACTCAGTAAGAAATTTATATTAGTCTGTTTGAAGTGAAGAAAGATTTCTCATCACTCAA 1560
DB |||||
QY 1561 CCTTATGACGAGGAAGGAGGCTTTTGAGAACCTTTACTTAGCAGAACCCATAT 1620
DB |||||
QY 1561 CCTTATGACGAGGAAGGAGGCTTTTGAGAACCTTTACTTAGCAGAACCCATAT 1620
DB |||||
QY 1621 TTTAGACACTTCCCTGCATTAATGACACAAACAATATGTTTGAACACTTTGTTGATCAAC 1680
DB |||||
QY 1621 TTTAGACACTTCCCTGCATTAATGACACAAACAATATGTTTGAACACTTTGTTGATCAAC 1679
DB |||||
QY 1681 CTCCAAACAGCACATTCAGGAGTTAAATATTTTTCATCAAAACATTTGGATTTTTCCTTA 1740
DB |||||
QY 1680 CTCCAAACAGCACATTCAGGAGTTAAATATTTTTCATCAAAACATTTGGATTTTTCCTTA 1739
DB |||||

QY 1741 ACGCTAGAGATTGCTCAAAATCTTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAAAGAT 1800
DB |||||
QY 1740 ACGCTAGAGATTGCTCAAAATCTTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAAAGAT 1799
DB |||||
QY 1801 TTCTCCCTGTTTATAGCAGCAAGACAAATTTAGCCATTTCTCTCTCAAACTTCACTAATGA 1860
DB |||||
QY 1800 TTCTCCCTGTTTATAGCAGCAAGACAAATTTAGCCATTTCTCTCTCAAACTTCACTAATGA 1859
DB |||||
QY 1861 TCACATTTCTTTCCAAAGGAACTCTAGAGACCCAAATGCCCGAGTTAAGAAACATCAAAA 1920
DB |||||
QY 1860 TCACATTTCTTTCCAAAGGAACTCTAGAGACCCAAATGCCCGAGTTAAGAAACATCAAAA 1919
DB |||||
QY 1921 CTAACCATCTGAAGAAACTTTCCCAAGTGTAAAGACTCTGCCATTTAAAACATTTACCGAGG 1980
DB |||||
QY 1920 CTAACCATCTGAAGAAACTTTCCCAAGTGTAAAGACTCTGCCATTTAAAACATTTACCGAGG 1979
DB |||||
QY 1981 GGACTCAAAACAGTCTTTCTTCTTTGTCGTGTTTCTTCTGCCAGACCAAGGCACCTGAGC 2040
DB |||||
QY 1980 GGACTCAAAACAGTCTTTCTTCTTTGTCGTGTTTCTTCTGCCAGACCAAGGCACCTGAGC 2039
DB |||||
QY 2041 ACAGTACTGTATACATAATTTTAAAGCACACTCCCTTTCCACTTTTGGTAAATACCAGAACTCT 2100
DB |||||
QY 2040 ACAGTACTGTATACATAATTTTAAAGCACACTCCCTTTCCACTTTGTTGTAATACCAGAACTCT 2099
DB |||||
QY 2101 AATTGGACCCCTGAAGCTTTAGGACTACAGCCATACAAATAGTAAACTCTGTCCACGA 2160
DB |||||
QY 2100 AATTGGACCCCTGAAGCTTTAGGACTACAGCCATACAAATAGTAAACTCTGTCCACGA 2159
DB |||||
QY 2161 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTATATAAANAATACCC 2220
DB |||||
QY 2160 TTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTATATAAANAATACCC 2219
DB |||||
QY 2221 CGGCAGACGAGTGGCTCAGCCCTGTAATCCAGCACTTTGGGAGGTGGGTGATCACC 2280
DB |||||
QY 2220 CGGCAGACGAGTGGCTCAGCCCTGTAATCCAGCACTTTGGGAGGTGGGTGATCACC 2279
DB |||||
QY 2281 TGAGTGGGAGTTTCGAGACCCAGCTGACAGCAATGGTGGAAACCCCATCTCTACTAAAA 2340
DB |||||
QY 2280 TGAGTGGGAGTTTCGAGACCCAGCTGACAGCAATGGTGGAAACCCCATCTCTACTAAAA 2339
DB |||||
QY 2341 ACACAAAAAATTTAGCCGGCGTGGTGGCACATGCTCTGTAAATCCAGCTACTCAGGAGGCT 2400
DB |||||
QY 2340 ACACAAAAAATTTAGCCGGCGTGGTGGCACATGCTCTGTAAATCCAGCTACTCAGGAGGCT 2399
DB |||||
QY 2401 GAGGCGGAGAAATTTGCTTGAACCCCGAAGGTGGAGGTGTTGCGGTGAGCTGAGATTGCAC 2460
DB |||||
QY 2400 GAGGCGGAGAAATTTGCTTGAACCCCGAAGGTGGAGGTGTTGCGGTGAGCTGAGATTGCAC 2459
DB |||||
QY 2461 TATTGCACTCCAGCTGGGCAACAGGAGTAAAACTCCCCCCCCACCC 2506
DB |||||
QY 2460 TATTGCACTCCAGCTGGGCAACAGGAGTAAAACTCCCCCCCCACCC 2505
DB |||||

RESULT 3

AAZ87583
ID AAZ87583 standard; DNA; 2088 BP.

XX AAZ87583;
AC AAZ87583;

XX 19-APR-2000 (first entry)
DT 19-APR-2000 (first entry)

XX Prostate disease marker UC Band #28.
DE Prostate disease marker UC Band #28.

XX Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; PCR primer; ss.

XX Homo sapiens.
OS Homo sapiens.

XX WO9964631-A1.
PN WO9964631-A1.

XX 16-DEC-1999.
PD 16-DEC-1999.

XX

PP	11-JUN-1999;	99WO-US013151.
XX		
PR	12-JUN-1998;	98US-00097199.
XX		
PA	(UROC-) UROCOR INC.	
XX		
PI	An G, O'hara SM, Ralph D, Veltri RW;	
XX		
DR	WPI; 2000-116557/10.	
DR	P-PsDB; AAY59295.	
XX		
PT	Novel RNA biomarkers for diagnosis, prognosis and management of prostate, breast and bladder cancer.	
PT		
XX		
PS	Claim 1; Page 182-183; 191pp; English.	
XX		
CC	The invention provides nucleic acid markers of prostate, breast and bladder cancer. The markers are indicators of malignant transformation of prostate, breast and bladder tissues and are diagnostic of the potential for metastatic spread of malignant prostate tumours. The nucleic acid can also be used as targets for therapeutic intervention in prostate cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The markers may be used to design specific probes and primers for the rapid analysis of prostate, bladder or breast biopsy samples. The probes and primers may also be used for in situ hybridization or in situ PCR detection and diagnosis. They may also be used to identify and isolate full length gene sequences from various DNA libraries. Antibodies against the polypeptide products of the markers can be used to treat prostate cancer, bladder cancer or breast cancer. The encoded proteins may be used to detect antibodies. The proteins and antibodies can be used in immunodetection methods for detecting or quantifying the cancers, and for clinical diagnosis of these cancers. The antibodies may also be used for radioimaging to quantify and localize the encoded proteins	
XX		
SQ	Sequence 2088 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 1 Other;	
Query Match 78.3%; Score 1961.6; DB 3; Length 2088;		
Best Local Similarity 99.5%; Pred. No. 0;		
Matches 1968; Conservative 0; Mismatches 10; Indels 0; Gaps 0;		
Qy	1	GACCTTAATATATCGAGTGGCTAATTGATGTATATAATTTACAAAATTTCTTCTTA
Dd	1	
Qy	61	TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCCATGAGGGCCTTCTTAAAGAACC
Dd	61	TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCCATGAGGGCCTTCTTAAAGAACC
Qy	121	AGAAATATGAGGATATGCACAATATTATTTCACATTTTACAGATCAGAAAAATGAGGCACA
Dd	121	AGAAATATGAGGATATGCACAATATTATTTCACATTTTACAGATCAGAAAAATGAGGCACA
Qy	181	GATTAAAGTAACCTCCCAAGGCTACACAGGATTTCTAGCTCCAGAAAATGTGCTTTACCAT
Dd	181	GATTAAAGTAACCTCCCAAGGCTACACAGGATTTCTAGCTCCAGAAAATGTGCTTTACCAT
Qy	241	TCTGCTACAAGGTATTTTCGAAAAAAGAAAAAGTAAAGGAAGTCAAAGGGCAACAGAGT
Dd	241	TCTGCTACAAGGTATTTTCGAAAAAAGAAAAAGTAAAGGAAGTCAAAGGGCAACAGAGT
Qy	301	TCATTGATTTATCCATGAAATGTTTCTCTCATTTATCTTCAGAACCATATTATTCA
Dd	301	TCATTGATTTATCCATGAAATGTTTCTCTCATTTATCTTCAGAACCATATTATTCA
Qy	361	TGACCATGAAAGGGTTCCTCAATGAAATGTTTCTCTCATTTATCTTCAGAACCATATTATTCA
Dd	361	TGACCATGAAAGGGTTCCTCAATGAAATGTTTCTCTCATTTATCTTCAGAACCATATTATTCA
Qy	421	CATTGACTTTTCAGTTAACTCAGACCCCTCAGGTCGTGGAATGCTGCTCTCTACTTATCCA
Dd	421	CATTGACTTTTCAGTTAACTCAGACCCCTCAGGTCGTGGAATGCTGCTCTCTACTTATCCA
Qy	481	AAACTATACATCCACAGATCATATAAACCTCTCAGCGCCCTGCTGCAAGGCTTTCCACAGAAA
Dd	481	AAACTATACATCCACAGATCATATAAACCTCTCAGCGCCCTGCTGCAAGGCTTTCCACAGAAA

481	AAAC	CTATACAT	CCACAGATCATATAA	CTCTCAGCCCTCTG	CTCAAGAGCCTTTCCAGAAA	540
541	ATAAAA	TGGTGGAAAAGG	CAATTCCTGCTACCAATG	ACTGTTTAAAGCCACGCAAGTAAC	600	
541	ATAAAA	TGGTGGAAAAGG	CAATTCCTGCTACCAATG	ACTGTTTAAAGCCACGCAAGTAAC	600	
601	TGAACCA	TTCCAACTTCAA	TTTATCTTATGAAAAGAA	TTTGATGTAGTAGGAGTTATTTTC	660	
601	TGAACCA	TTCCAACTTCAA	TTTATCTTATGAAAAGAA	TTTGATGTAGTAGGAGTTATTTTC	660	
661	AAATTC	TAAAAATACAAACCC	ATGTTGATCTTTCTCAATCTTGAA	CTCATAGATATTAATCT	720	
661	AAATTC	TAAAAATACAAACCC	ATGTTGATCTTTCTCAATCTTGAA	CTCATAGATATTAATCT	720	
721	ATTATCT	CAATTTAGTTTGTTATTTATCTAGTGGCC	ATAAATAAACAATCAATCAATGTT	780		
721	ATTATCT	CAATTTAGTTTGTTATTTATCTAGTGGCC	ATAAATAAACAATCAATCAATGTT	780		
781	TCGTCTCT	CCATTAGTCAATAA	CTAAACGAGCAATTAGTAAGCCATGTGCCAGAT	840		
781	TCGTCTCT	CCATTAGTCAATAA	CTAAACGAGCAATTAGTAAGCCATGTGCCAGAT	840		
841	GCTCCG	TAGGCAACGAGGGAT	TAATAAATAATCTTAGTATACCACTAATTTTCGTT	900		
841	GCTCCG	TAGGCAACGAGGGAT	TAATAAATAATCTTAGTATACCACTAATTTTCGTT	900		
901	AGTAACT	TAGTGAATGTTCTCAAGT	CTAGTCAAGAGTCAAGAGTTGAGGAGACATTAACAATGT	960		
901	AGTAACT	TAGTGAATGTTCTCAAGT	CTAGTCAAGAGTCAAGAGTTGAGGAGACATTAACAATGT	960		
961	GTAATG	AAAAACCAAGGAAAGT	GTAAAGTGTGGGATAAGTGGGACTAGTGTA	1020		
961	GTAATG	AAAAACCAAGGAAAGT	GTAAAGTGTGGGATAAGTGGGACTAGTGTA	1020		
1021	TAAT	TGATTTCTGACTCTAT	CAATGGCTCTCCAAAACAGATGTGTGTTTTCTTTGGTTTT	1080		
1021	TAAT	TGATTTCTGACTCTAT	CAATGGCTCTCCAAAACAGATGTGTGTTTTCTTTGGTTTT	1080		
1081	GTTTTCT	TTCACTATGGGATCTCTG	TGCCACGACAGTGCCTGACACATAGAAAAACAATC	1140		
1081	GTTTTCT	TTCACTATGGGATCTCTG	TGCCACGACAGTGCCTGACACATAGAAAAACAATC	1140		
1141	AAAT	TTTGCTGAAATAAATG	ATTAATAATCAAGAAATCTTCCCAATCTGTGTTGGATCTAT	1200		
1141	AAAT	TTTGCTGAAATAAATG	ATTAATAATCAAGAAATCTTCCCAATCTGTGTTGGATCTAT	1200		
1201	AGAA	CATCCAGATAGTGA	TGAGGCGCTCTGCATTTATATGCGCTTAAATTAAGATTTAT	1260		
1201	AGAA	CATCCAGATAGTGA	TGAGGCGCTCTGCATTTATATGCGCTTAAATTAAGATTTAT	1260		
1261	GTGAAA	AGTTTTAAAGACAC	TTTAGTGTAGGTGATTTTGAATATAGTAAACACCTTGGAAA	1320		
1261	GTGAAA	AGTTTTAAAGACAC	TTTAGTGTAGGTGATTTTGAATATAGTAAACACCTTGGAAA	1320		
1321	TGGTGG	TGCTTTAAAAAGATATTA	ATAGATATATGAAAAATCTCCATCTCAAAAATAATG	1380		
1321	TGGTGG	TGCTTTAAAAAGATATTA	ATAGATATATGAAAAATCTCCATCTCAAAAATAATG	1380		
1381	CATAA	ACTATTTAAAGGAAAT	CAATCTCCAGGCTTTCAATGTGTTGTTCAATCTTTT	1440		
1381	CATAA	ACTATTTAAAGGAAAT	CAATCTCCAGGCTTTCAATGTGTTGTTCAATCTTTT	1440		
1441	CATAT	TTTTTACCATCTC	GTGAAGGCAGTCATATCAAAAGGGTAAAGAAAGATGGGAGGA	1500		
1441	CATAT	TTTTTACCATCTC	GTGAAGGCAGTCATATCAAAAGGGTAAAGAAAGATGGGAGGA	1500		
1501	AACT	CAGTAAAGAAATTA	TATTAATGCTGTTGCAAAAGTAGAAAAAGATTTCTCATC	1560		
1501	AACT	CAGTAAAGAAATTA	TATTAATGCTGTTGCAAAAGTAGAAAAAGATTTCTCATC	1560		
1561	CCTT	TATGAGCAGGAAGG	GAAGGCTGTTTGTGAAACCAATTTTACTTAGCAGAAACCAATAT	1620		

Db 1561 CCTATTGACGAGGAGGAGGCTGTTTGTGAAACCATTTACTTAGCAGAACCAATAT 1620
QY 1621 TTTAGACACTTCCTTCATTAAGTCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
Db 1621 TTTAGACACTTCCTTCATTAAGTCACAAACAATATGTTTGCAAACTTGTTRGATCAAC 1680
QY 1681 CTCCTAACACGACACATTCAGGAGTTAAATATATTTTTCATCAAAACATTTGATTTTCCTTA 1740
Db 1681 CTCCTAACACGACACATTCAGGAGTTAAATATATTTTTCATCAAAACATTTGATTTTCCTTA 1740
QY 1741 AGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT 1800
Db 1741 AGCTAGAGATTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT 1800
QY 1801 TTCTCCCTGTTTAAAGCAGACAGACAAATTAGCCATTTTCACTCTCAAACTTCACTAATGA 1860
Db 1801 TTCTCCCTGTTTAAAGCAGACAGACAAATTAGCCATTTTCACTCTCAAACTTCACTAATGA 1860
QY 1861 TCACATTTCTTCCAAAGGAACTCTAGAGACACAAATGCCCCGAGTTAAGACATCAAAA 1920
Db 1861 TCACATTTCTTCCAAAGGAACTCTAGAGACACAAATGCCCCGAGTTAAGACATCAAAA 1920
QY 1921 CTAACCATCTGAAGAACTTCCCAAGTGAAGACTCTGCCATTAAACACATTACCGAGA 1978
Db 1921 CTAACCATCTGAAGAACTTCCCAAGTGAAGACTCTGCCATTAAACACATATAA 1978

RESULT 4
AAS04000
ID AAS04000 standard; cDNA; 2087 BP.
AC AAS04000;
XX
29-AUG-2001 (first entry)
XX Biomarker UC band 28 #2, used in diagnosis and prognosis of cancer.
XX
XX Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 99..506
FT CDS /*tag= a
FT /product= "Prostate cancer marker protein"
XX
XX US6218529-B1.
XX
XX 17-APR-2001.
XX
XX 12-JUN-1998; 98US-00097199.
XX
XX 31-JUL-1995; 95US-0001655P.
XX 11-JAN-1996; 96US-0013611P.
XX 31-JUL-1996; 96US-00692787.
XX
XX (UROC-) UROCOR INC.
XX
XX An G, O'hara SM, Ralph D, Veltri R;
XX
XX WPI; 2001-289849/30.
DR P-PSDB; AAU02174.
XX
XX New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
XX Claim 2; Col 117-121; 78pp; English.
XX
XX The sequence represents nucleic acid biomarker UC band 28 #2, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that

CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ Sequence 2087 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 0 Other;
Query Match 77.8%; Score 1950; DB 4; Length 2087;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 1 GACCTTAAATATATCGAGGTGCTAAATTTGATGATGATATAATAATTTACAAAATTTATTTCTTA 60
Db 1 GACCTTAAATATATCGAGGTGCTAAATTTGATGATGATATAATAATTTACAAAATTTATTTCTTA 60
QY 61 TTGCTACAGAGCTACAATTTCAATTTTACAGTAGGCCACCATGAGGCTTCTTTAAGGAACC 120
Db 61 TTGCTACAGAGCTACAATTTCAATTTTACAGTAGGCCACCATGAGGCTTCTTTAAGGAACC 120
QY 121 AGAAATATGAGGATATGCACAATATTTATTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
Db 121 AGAAATATGAGGATATGCACAATATTTATTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
QY 181 GATTAAAGTAACTTCCCAAGGCTACAGGATCTTAGCTCCAGAACTGTCTTTACCAT 240
Db 181 GATTAAAGTAACTTCCCAAGGCTACAGGATCTTAGCTCCAGGCAATTTAGCTCTTTACCAT 240
QY 241 TCTGCTACAAGTATTTTCGAAAAAGAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 300
Db 241 TCTGCTACAAGTATTTTCGAAAAAGAAAGTAAAGAAAGTCAAAAGGCAACAGAGT 300
QY 301 TCATTGATTATTTCCATAGACAGTCAACCATGCAATTTCTCACACCTTTGCAGACACACT 360
Db 301 TCATTGATTATTTCCATAGACAGTCAACCATGCAATTTCTCACACCTTTGCAGACACACT 360
QY 361 TGACCATGAAAGGTTTCTCAATGAAATGTTCTCATTTATCTTCAGAAGCCATATTTTCA 420
Db 361 TGACCATGAAAGGTTTCTCAATGAAATGTTCTCATTTATCTTCAGAAGCCATATTTTCA 420
QY 421 CATTGACCTTTGCAGTTAACTCAGACCTAGCTGGAATGCTCTCTCTACTTTATCCA 480
Db 421 CATTGACCTTTGCAGTTAACTCAGACCTAGCTGGAATGCTCTCTCTACTTTATCCA 480
QY 481 AAACCTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGGCTTTTCCAGAAAA 540
Db 481 AAACCTATACATCCACAGATCATATAAACTCTCAGCCCTGCTGCAAGGCTTTTCCAGAAAA 540
QY 541 ATAAAAATGGTTGAAAAGGCAATTCGTCTACCAATGACTGTTTAAAGCCCGCAAGTAAC 600
Db 541 ATAAAAATGGTTGAAAAGGCAATTCGTCTACCAATGACTGTTTAAAGCCCGCAAGTAAC 600
QY 601 TGAACCATTTCCAACTTCAATTTTACTTATGAAAAGAAATTTGATGATGAGGAGTTATTTC 660
Db 601 TGAACCATTTCCAACTTCAATTTTACTTATGAAAAGAAATTTGATGATGAGGAGTTATTTC 660
QY 661 AATTCTAAAAATACAAACCCATGTTGATCTTCTCAATCTTGAACCTCATAGATTATTATCT 720
Db 661 AATTCTAAAAATACAAACCCATGTTGATCTTCTCAATCTTGAACCTCATAGATTATTATCT 720
QY 721 ATTATCTCAATTTAGTTTGTATTATCTTAGTGGGCCATTAAAAAATACCACATGTGT 780
Db 721 ATTATCTCAATTTAGTTTGTATTATCTTAGTGGGCCATTAAAAAATACCACATGTGT 780
QY 781 TCTGCTCTCCATTAGTCAATAAATCTAAACTAACGAGCAATTAGTAGGCATGTGCCAGAT 840
Db 781 TCTGCTCTCCATTAGTCAATAAATCTAAACTAACGAGCAATTAGTAGGCATGTGCCAGAT 840

Qy	841	GCTCCGCTAGGCACACAGAGGATATAAAACAATACTTTATAGTATACCACTAAATTTTCGCTT	900
Db	841	GCTCCGCTAGGCACACAGAGGATATAAAACAATACTTTATAGTATACCACTAAATTTTCGCTT	900
Qy	901	AGTAACTAGTGAATGTTCAAGTTCATCCCTGAGTCAAGAGTTGAGGAGACATATACAATGT	960
Db	901	AGTAACTAGTGAATGTTCAAGTTCATCCCTGAGTCAAGAGTTGAGGAGACATATACAATGT	960
Qy	961	GTAATGAAACCAAGGAAAGTGAACTTTGGATAAGTGGGACCTAGTGTATTTATATATTT	1020
Db	961	GTAATGAAACCAAGGAAAGTGAACTTTGGATAAGTGGGACCTAGTGTATTTATATATTT	1020
Qy	1021	TAATTTGATTTCTGACTCTATCATTTGGCGCTCCAAACACAGATTTGTTTTCTTTGGTTTT	1080
Db	1021	TAATTTGATTTCTGACTCTATCATTTGGCGCTCCAAACACAGATTTGTTTTCTTTGGTTTT	1080
Qy	1081	GTTTTCTTCTACTATGGGATCTTTCTGTGCCAGACAGTGTCTGACACATAGAAAAACAATC	1140
Db	1081	GTTTTCTTCTACTATGGGATCTTTCTGTGCCAGACAGTGTCTGACACATAGAAAAACAATC	1140
Qy	1141	AATATTTTGTCTGAATTAATGATTTAAAAAATCAGAGAACTTTTCCCATTCTGTTTGGATCTAT	1200
Db	1141	AATATTTTGTCTGAATTAATGATTTAAAAAATCAGAGAACTTTTCCCATTCTGTTTGGATCTAT	1200
Qy	1201	AGAACTCCAGAGTAAGTATGAGGGGCTCTGCATTTATATGCGCTTAAATTAAGATTTAT	1260
Db	1201	AGAACTCCAGAGTAAGTATGAGGGGCTCTGCATTTATATGCGCTTAAATTAAGATTTAT	1260
Qy	1261	GTGAGAAAGTTTTAAAGACACTTAGTAGAGTGATTTTTGAAATATAGTAACACTTGGAAA	1320
Db	1261	GTGAGAAAGTTTTAAAGACACTTAGTAGAGTGATTTTTGAAATATAGTAACACTTGGAAA	1320
Qy	1321	TGGTGGTGCTTTAAAAAGATTAATAAGATAAATAGAAAAATCTCCATCTCAAAAAATAATG	1380
Db	1321	TGGTGGTGCTTTAAAAAGATTAATAAGATAAATAAGAAAAATCTCCATCTCAAAAAATAATG	1380
Qy	1381	CATAAACTATTTAAAGGAAATACACTCTCCAGCGCTTTCAATGTTTGTTCATTACTTTTT	1440
Db	1381	CATAAACTATTTAAAGGAAATACACTCTCCAGCGCTTTCAATGTTTGTTCATTACTTTTT	1440
Qy	1441	CATATATTTTACCATCTGCTGAAGGCAGTCATATCAAGGGTAAAGAAAGATGGGAGGA	1500
Db	1441	CATATATTTTACCATCTGCTGAAGGCAGTCATATCAAGGGTAAAGAAAGATGGGAGGA	1500
Qy	1501	AAACTCAGTAAGAAATTAATTTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA	1560
Db	1501	AAACTCAGTAAGAAATTAATTTAGTCTGTTTGCAAAGTAGAAAAAGATTCTCATCACTCAA	1560
Qy	1561	CTTTATGAGCAGGAAGGGAAGCGCTGTTTGAAGAACCATTTACTTTAGCAGAACCCACATAT	1620
Db	1561	CTTTATGAGCAGGAAGGGAAGCGCTGTTTGAAGAACCCATTTACTTTAGCAGAACCCACATAT	1620
Qy	1621	TTTATAGACATTTCCCTGCATTTAACTGCAACAAATATGTTTGCMAACTTGTGTGATCAAC	1680
Db	1621	TTTATAGACATTTCCCTGCATTTAACTGCAACAAATATGTTTGCMAACTTGTGTGATCAAC	1680
Qy	1681	CTCCAAACAAACGACACATTTCAGGAGTTAAATATTTTTTCATCAAAACATTGGATTTTTCTTTA	1740
Db	1680	CTCCAAACAAACGACACATTTCAGGAGTTAAATATTTTTTCATCAAAACATTGGATTTTTCTTTA	1739
Qy	1741	ACGCTAGAGATTTGTACAAATCTTTCTGAAGGGTCTCAATGGCTTTCAGGCTTAAGAAGAGAT	1800
Db	1740	ACGCTAGAGATTTGTACAAATCTTTCTGAAGGGTCTCAATGGCTTTCAGGCTTAAGAAGAGAT	1799
Qy	1801	TTTCTCCCTGTTATAAGCAGCAAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATGA	1860
Db	1800	TTTCTCCCTGTTATAAGCAGCAAGACAAATTTAGCCATTTTCACTCTCAAACTTCACTAATGA	1859
Qy	1861	TCACATTTCTTTCCAAAAGGAACTCTAGACAGCAAAATGCCCGGAGTTTAAAGACATCAAAA	1920
Db	1860	TCACATTTCTTTCCAAAAGGAACTCTAGACAGCAAAATGCCCGGAGTTTAAAGACATCAAAA	1919
Qy	1921	CTAAACCATCTGAAGAACTTTCCCAAGTGTGAAGACTCTGCGCAATTTAAAAACATTAACCGAGA	1978

Db	1920	CTAACCATCTGAAGAAATCTTCCCAAGTGTAAAGACTCTGCGTCGACGCAACACATATAA	1977
RESULT 5			
AAV16883			
ID	AAV16883	standard; DNA; 757 BP.	
XX	AAV16883;		
XX	07-AUG-1998	(first entry)	
XX	Human prostate cancer marker UC Band #28.		
XX	Prostate cancer; human; marker; diagnosis; treatment; probe; ss.		
OS	Homo sapiens.		
PN	W05804689-A1.		
XX	05-FEB-1998.		
XX	31-JUL-1996;	96WO-US012516.	
XX	31-JUL-1996;	96WO-US012516.	
XX	(UROC-) UROCOR INC.		
XX	Veltri R, Ohara SM, An G, Ralph D;		
XX	WPI; 1998-130681/12.		
XX	Human prostate cancer marker - useful for detection and treatment of		
XX	human prostate cancer.		
XX	Claim 1; Page 133-134; 229pp; English.		
XX	This represents a marker sequence for human prostate cancer. Isolated		
XX	nucleic acid segments shown in AAV16881 to AAV16885, AAV16890 to		
XX	AAV16903, AAV26351 and AAV26352 which can act as human prostate cancer		
XX	markers are provided in the specification. It also provides methods for		
XX	identifying markers for human prostate cancer and for detection of		
XX	prostate cancer cells. The markers can be identified by amplifying human		
XX	prostate RNA to provide nucleic acid amplification products, separating		
XX	the products and identifying those RNA that are differentially expressed		
XX	between human prostate cancers versus normal or benign human prostate.		
XX	Prostate cancer cells in a sample can be detected by detecting a nucleic		
XX	acid in a sample, the nucleic acid being a prostate cancer marker.		
XX	Primers and probes derived from this marker can be used for the detection		
XX	of prostate cancer cells in a sample. Antibodies against the protein		
XX	encoded by the marker nucleic acid fragments, inhibitors of the protein		
XX	and oligonucleotides antisense to the markers can be used in the		
XX	treatment of prostate cancer. The antibodies can also be used for the		
XX	diagnosis of human prostate cancer		
XX	Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;		
Query Match	25.0%;	Score 626.2; DB 2; Length 757;	
Best Local Similarity	94.6%;	Pred. No. 1.3e-135;	
Matches	716; Conservative	0; Mismatches 28; Indels 13; Gaps 6;	
QY	1686	ACACGACACATTCAGGAGTTAAATATTTTTCATCAACATTCGATTTTTCCTTAAGCT	1745
Db	1	ACACGACACATTCAGGAGTTAAATATTTTTCATCAACATTCGATTTTTCCTTAAGCT	60
QY	1746	AGAGATTGCTACAAATCTTCTTGAAGGCTCTCAATGGCTTCAGGCTAAGAGAGATTTC	1805
Db	61	AGAGATTGCTACAAATCTTCTTGAAGGCTCTCAATGGCTTCAGGCTAAGAGAGATTTC	120
QY	1806	CTGTGTATAGCAGCAAGACAAATTAGCCATTTCCTCAACTCAAACTTCACATGATCACA	1865
Db	121	CTGTGTATAGCAGCAAGACAAATTAGCCATTTCCTCAACTCAAACTTCACATGATCACA	180

1866	Qy	TTCTTTCCAAAAGGAACCTCTAGAAAGACCAAAATGCCGAGTTTAAGAAACATCAAAAACCTAAC	1925
181	Db	TTCTTTCCAAAAGGAACCTCTAGAAAGACCAAAATGCCGAGTTAAGAAACATCAAAAACCTAAC	240
1926	Qy	CATCTGAAGAAACCTTCCCAAGCTAGAGACTCGCCATTAAACCATTTACCCAGAGGGGACT	1985
241	Db	CATCTGAAGAAACCTTCCCAAGCTAGAGACTCGCCATTAAACCATTTACCCAGAGGGGACT	300
1986	Qy	CAAAACAGTCTTTT--CTTCCCTTTGTCGTGTTTC--TTGCTCCAGACCAAGG-----CACT	2036
301	Db	CAAAACAGTCTTTTCCCTTTGTCGTGTTTC--TTGCTCCAGACCAAGGCTGGCG	360
2037	Qy	GACGACAGTACTGATACATAAATTTAAAGCA-CATCCCTTCCACTTTTGTA--TACCA	2093
361	Db	GACAGTACTTGTATACATAAATTTAAAGCACCACTCCCTTCCACTTTTGTAAATACCCA	420
2094	Qy	GAACCTCAATTGGACCAACCTGAAGCTTAGGA-CTACAGCGCATACAAATAGTAACTCT	2152
421	Db	GAACCTCAATTGGACCAACCTGAAGCTTAGGACCTACAGCGCATACAAATAGTAACTCT	480
2153	Qy	GTCCAGGATTCATCATCTCTGTATTTTCTATAGATGTTTACTAGGCGTTGTGTATATAA	2212
481	Db	GTCCAGGATTCATCATCTCTGTATTTTCTATAGATGTTTACTAGGCGTTGTGTATATAA	540
2213	Qy	AAATACCCCGGCAGGCA CGGTGGCTCA CGCCTGTATATCCAGACATTTGGGAGGTGGGT	2272
541	Db	AAATACCCCGGCAGGCA CGGTGGCTCA CGCCTGTATATCCAGACATTTGGGAGGTGGGT	600
2273	Qy	GGATCACCTGAGGTCGGGAGTTTGAGACACAGCCTGACCAGCATGGTGGAAACCCCATCTC	2332
601	Db	GGATCACCTGAGGTCGGGAGTTTGAGACACAGCCTGACCAGCATGGTGGAAACCCCATCTC	660
2333	Qy	TACTAAAAACACAAAAAATTAGCGGGCGTGGTGACATGCTCTGTATATCCAGCTACTC	2392
661	Db	TACTAAAAACACAAAAAATTAGCGGGCGTGGTGACATGCTCTGTATATCCAGCTACTC	720
2393	Qy	AGGAGGCTAGGCGGAGAAATTGCTTGAAACCCGGAAGG	2429
721	Db	AGGAGGCTAGGCGGAGAAATTGCTTGAAACCCGGAAGG	757

RESULT 6

AAZ26018	
ID	AAZ26018 standard; DNA; 757 BP.
XX	XX
XX	AAZ26018;
XX	XX
DT	20-MAY-1999 (first entry)
XX	XX
DE	Prostate disease marker gene fragment UC Band #28.
XX	XX
KW	Prostate cancer; benign prostatic hyperplasia; marker gene; tumour;
XX	differentiation; Reverse Transcription Polymerase Chain Reaction;
KW	diagnostic; progression; cancer; metastasis; human; RT-PCR; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	US5882864-A.
XX	XX
PD	16-MAR-1999.
XX	XX
PF	31-JUL-1996; 96US-00692787.
XX	XX
PR	31-JUL-1995; 95US-0001655P.
XX	XX
PA	(UROC-) UROCOR INC.
XX	XX
PI	Veltri R, Ralph D, An G, O'hara SM;
XX	XX
DR	WPI; 1999-214055/18.
XX	XX
PT	Diagnosing prostate cancer and benign prostatic hyperplasia cells - using
PT	oligonucleotide probes specific for marker genes associated with tumor

Db 721 AGGAGGCTGAGCGGAGAAATTGCTTGAACCCGAAGG 757

RESULT 7

AAZ87503

ID AAZ87503 standard; cDNA; 757 BP.

AC AAZ87503;

XX

DT 19-APR-2000 (first entry)

XX

DE Prostate, breast and bladder cancers detecting biomarker UC Band #28.

XX

KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
KW diagnosis; ss.

XX

OS Homo sapiens.

XX

PN WO9964631-A1.

XX

PD 16-DEC-1999.

XX

PF 11-JUN-1999; 99WO-US013151.

XX

PR 12-JUN-1998; 98US-00097199.

XX

PA (UROC-) UROCOR INC.

XX

PI An G, O'hara SM, Ralph D, Veltri RW;

XX

DR WPI; 2000-116557/10.

XX

PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.

XX

PS Claim 1; Page 165; 191pp; English.

XX

CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins

XX

SQ Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

XX

Query Match 25.0%; Score 626.2; DB 3; Length 757;

XX

Best Local Similarity 94.6%; Pred. No. 1.3e-135;

XX

Matches 716; Conservative 0; Mismatches 28; Indels 13; Gaps 6;

XX

QY 1686 ACAACGACATTCAGGAGTAAATATTTTCATCAACATTTGGATTTTCTTAAACGCT 1745

Db

1 ACAACGACATTCAGGAGTAAATATTTTCATCAACATTTGGATTTTCTTAAACGCT 60

QY

1746 AGAGATTGCTACAAATCTTCTGAGGGTCTCAATGGCTTCAGGCTTAAGAGAGATTCTC 1805

Db

61 AGAGATTGCTACAAATCTTCTGAGGGTCTCAATGGCTTCAGGCTTAAGAGAGATTCTC 120

QY

1806 CTGTGTTAAGCAGCAGCAAAATAGGCAATTTTCACTCTCAAACTTCACTAATGATCACA 1865

Db

121 CTGTGTTAAGCAGCAGCAAAATAGGCAATTTTCACTCTCAAACTTCACTAATGATCACA 180

QY 1866 TTCTTTCCAAAAGGAACTCTTAGAAGACCAAAATGCCCGAGTTTAAAGACATCAAAACTAAC 1925
Db 181 TTCTTTCCAAAAGGAACTCTTAGAAGACCAAAATGCCCGAGTTTAAAGACATCAAAACTAAC 240
QY 1926 CATCTGAGAGAACTTCCCAAGGTGAAGACTCTGCCATTAAACANTTCCGAGAGGGGACT 1985
Db 241 CATCTGAGAGAACTTCCCAAGGTGAAGACTCTGCCATTAAACANTTCCGAGAGGGGACT 300
QY 1986 CAAACAGTCTTT--CTTCCTTTGTGCTGTGTTTC--TTGCTCCCAAGCAAGG-----CACT 2036
Db 301 CAAACAGTCTTTCTTCCCTTTGTGCTGTGTTTC--TTGCTCCCAAGCAAGGCACTTTGGCG 360
QY 2037 GACGACAGTACTGATACATAATTTAAAGCA--CACTCCCTTCCACTTTGGTTAA--TACCA 2093
Db 361 GACAGTACTTGATACATAATTTAAAGCAAGCACTCCCTTCCCACTTTGTAAATACCCA 420
QY 2094 GAACTCTTAATTGGACCACTTGAAGCTTAGGA--CTACCAGCCATCAAAATAGTAACTCT 2152
Db 421 GAACTCTTAATTGGACCACTTGAAGCTTAGGA--CTACCAGCCATCAAAATAGTAACTCT 480
QY 2153 GTCCACGATTCACTCATCTGTGTTATTTCTATAGATGTTTACTAGGCGTTTCTTATATAA 2212
Db 481 GTCCACGATTCACTCATCTGTGTTATTTCTATAGATGTTTACTAGGCGTTTCTTATATAA 540
QY 2213 AAATACCCCGCCAGGACGGTGGCTCAGCGCTGTAATCCCAAGCACTTTGGAGAGTGGGT 2272
Db 541 AAATACCCCGCCAGGACGGTGGCTCAGCGCTGTAATCCCAAGCACTTTGGAGAGTGGGT 600
QY 2273 GGATCACCCTGAGGTCGGGAGTTCGAGACCCGCTGACCAAGCATGGTGGAAACCCCATCTC 2332
Db 601 GGATCACCCTGAGGTCGGGAGTTCGAGACCCGCTGACCAAGCATGGTGGAAACCCCATCTC 660
QY 2333 TACTAAAAACACAAAAAATTAGCCGGGCTGTGGCAGCATGCTGTAATCCCAAGTACTC 2392
Db 661 TACTAAAAACACAAAAAATTAGCCGGGCTGTGGCAGCATGCTGTAATCCCAAGTACTC 720
QY 2393 AGGAGGCTGAGCGGAGAAATTGCTTTGAACCCGGAAGG 2429
Db 721 AGGAGGCTGAGCGGAGAAATTGCTTTGAACCCGGAAGG 757

RESULT 8

AA503722

ID AA503722 standard; cDNA; 757 BP.

XX

AC AA503722;

XX

DT 29-AUG-2001 (first entry)

XX

DE Biomarker UC band 28 #1, used in diagnosis and prognosis of cancer.

XX

KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;

XX

KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.

XX

OS Homo sapiens.

XX

PN US6218529-B1.

XX

PD 17-APR-2001.

XX

PF 12-JUN-1998; 98US-00097199.

XX

PR 31-JUL-1995; 95US-0001655P.

XX

PR 11-JAN-1996; 96US-0013611P.

XX

PR 31-JUL-1996; 96US-00692787.

XX

PA (UROC-) UROCOR INC.

XX

PI An G, O'hara SM, Ralph D, Veltri R;

XX

DR WPI; 2001-289849/30.

PT New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
PS
XX Claim 2; Col 79; 78pp; English.

CC The sequence represents nucleic acid biomarker UC band 28 #1, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that
CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX
SQ Sequence 757 BP; 228 A; 190 C; 147 G; 192 T; 0 U; 0 Other;

Query Match 25.0%; Score 626.2; DB 4; Length 757;
Best Local Similarity 94.6%; Pred. No. 1.3e-135;
Matches 716; Conservative 0; Mismatches 28; Indels 13; Gaps 6;

QY 1686 ACAACGACATTCAGGAGCTAAATATTTTTCATCAAAATTTGGATTTTTCCTTAACGCT 1745
DB 1 ACACGACATTCAGGAGCTAAATATTTATCATCAAAATTTGGATTTTTCCTTAACGCT 60

QY 1746 AGAGATTGTCACAAATCTTCTGAAGGCTCAATAGCTTTCAGGCTTAAGAAGATTTCTC 1805
DB 61 AGAGATTGTCACAAATCTTCTGAAGGCTCAATAGCTTTCAGGCTTAAGAAGATTTCTC 120

QY 1806 CTTGTTATAGCAGCAGCAATATAGCCATTTCACTCTCAAACTTCACTATGATCACA 1865
DB 121 CTTGTTATAGCAGCAGCAATATAGCCATTTCACTCTCAAACTTCACTATGATCACA 180

QY 1866 TTCTTTCCAAAAGAACTCTAGAAGACCAATGCCCCGAGTTAAGAAATCAAACTAAC 1925
DB 181 TTCTTTCCAAAAGAACTCTAGAAGACCAATGCCCCGAGTTAAGAAATCAAACTAAC 240

QY 1926 CATCTGAAGAAATCTCCCAAGTGTAAGACTCTGCCATTTAAACATTTACCGAGGGGACT 1985
DB 241 CATCTGAAGAAATCTCCCAAGTGTAAGACTCTGCCATTTAAACATTTACCGAGGGGACT 300

QY 1986 CAAACAGTCTTTTCTTCCCTTTTGTGCTGTTTC-TTGCTCCAGACCAAGG-----CACT 2036
DB 301 CAAACAGTCTTTTCTTCCCTTTTGTGCTGTTTC-TTGCTCCAGACCAAGGCACTTGGG 360

QY 2037 GACGACAGTACTGATACATAATTTAAAGCA-CACTCCCTTCCACTTTGGTAA--TACCA 2093
DB 361 GACGACAGTACTGATACATAATTTAAAGCACTCCCTTCCACTTTGGTAAATACCA 420

QY 2094 GAACTCTAATAGCACCACTGAGCTTAGGA-CTACACGACCATCAAAATAGTAACTCT 2152
DB 421 GAACTCTAATAGCACCACTGAGCTTAGGA-CTACACGACCATCAAAATAGTAACTCT 480

QY 2153 GTCCAGATTCATCATCTGTATTTTCTATAGATGTTTACTAGCGTTGTATATAA 2212
DB 481 GTCCAGATTCATCATCTGTATTTTCTATAGATGTTTACTAGCGTTGTATATAA 540

QY 2213 AAATACCCCGGCGAGCAGCGTGGCTCAGCCCTGTATCCAGCACTTTGGGAGGTGGGT 2272
DB 541 AAATACCCCGGCGAGCAGCGTGGCTCAGCCCTGTATCCAGCACTTTGGGAGGTGGGT 600

QY 2273 GGATCACTGAGGTGGGAGTTTCAGACCAAGCTGACACAGATGTTGGAAACCCCACTCTC 2332
DB 601 GGATCACTGAGGTGGGAGTTTCAGACCAAGCTGACACAGATGTTGGAAACCCCACTCTC 660

QY 2333 TACTTAAACACAAAATTTAGCCGGGCTGGTGGCAGATGCTGTAAATCCAGCTACTC 2392
DB 661 TACTTAAACACAAAATTTAGCCGGGCTGGTGGCAGATGCTGTAAATCCAGCTACTC 720

QY 2393 AGGAGGCTGAGCGGAGAAATTGCTTGAACCCGGAAGG 2429
DB 721 AGGAGGCTGAGCGGAGAAATTGCTTGAACCCGGAAGG 757

RESULT 9
ADQ59380/c
ID ADQ59380 standard; DNA; 299598 BP.
XX
AC ADQ59380;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer-associated (CA) gene sequence SEQ ID NO:16.
XX
KW human; cancer-associated gene; cancer-associated protein; cytostatic;
KW gene therapy; vaccine; tyrosine kinase antagonist;
KW G-protein coupled receptor antagonist; cancer; lymphoma; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004058288-A1.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040082.
XX
PR 17-DEC-2002; 2002US-00322696.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
WI WIPI; 2004-543349/52.
XX
P-PsDB; ADQ59382.
XX
PT New cancer-associated nucleic acid for diagnosing, preventing or treating
PT cancer (e.g. lymphoma) or for screening agents that may be used for
PT treating or preventing cancer.
XX
PS Claim 16; SEQ ID NO 16; 143pp; English.

The present invention describes human cancer-associated (CA) nucleotide sequences (1). Also described: (1) an expression vector comprising (1); (2) a host cell comprising (1) or the expression vector; (3) a microarray for detecting a CA nucleic acid; (4) an isolated polypeptide encoded within an open reading frame of a CA sequence; (5) an isolated antibody, or its antigen binding fragment, that binds to the above polypeptide; (6) a hybridoma that produces the monoclonal antibody described above; (7) a pharmaceutical composition comprising the antibody and a pharmaceutical excipient; (8) a kit for detecting or diagnosing cancer cells, comprising the above (monoclonal) antibody or polynucleotide that selectively hybridises to any of the polynucleotide sequences mentioned above; (9) methods for diagnosing cancer or for detecting the presence or absence of cancer cells in an individual; (10) a method for inhibiting growth of cancer cells in an individual; (11) a method for delivering a therapeutic agent to cancer cells in an individual; (12) an electronic library comprising the polynucleotide or polypeptide, or their fragments, mentioned above; (13) a method of screening for anticancer activity; (14) methods for detecting cancer associated with expression of a polypeptide or the presence of the antibody in a test cell or serum sample; (15) a method for screening for a bioactive agent capable of modulating the activity of a CA protein encoded by the above nucleic acid molecule; and (16) a method for treating cancers. (1) has cytostatic activity, and can be used in gene therapy, in vaccines, as a tyrosine kinase antagonist, and as a G-protein coupled receptor antagonist. The compositions and methods of the present invention can be used for diagnosing, preventing and treating cancer, especially lymphomas. They may also be used in screening for agents that may be used for treating or preventing cancer. The present sequence represents a human CA gene sequence, which is given in the exemplification of the present invention. Note: The sequence data for this patent did not form part of the printed specification, but was

```
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 299598 BP; 92218 A; 54563 C; 55801 G; 95585 T; 0 U; 1431 Other;
Query Match 7.6%; Score 190.6; DB 12; Length 299598;
Best Local Similarity 78.4%; Pred. No. 2.1e-33;
Matches 269; Conservative 0; Mismatches 64; Indels 10; Gaps 3;
QY 2168 ATCTGTGATTTTCTATAGATTTTACTAGGCGTTTGTATATAAAATACCCGCCAG 2227
DB 241337 ATGTATTTTATCATATATTTATATGAGACTTAAATATATATTTAGGCCAG 241278
QY 2228 GCACGCTGCTCAGCGCTGTAATCCAGCACTTTGGAGGT-----GGTGTGATCACCT 2281
DB 241277 GTGTGTGGCTCAGCGCTGTAATCCAGCACTTTGGAGGCTGAGAGGTTGGATCACCT 241218
QY 2282 GAGCTCGGAGTTTCGAGACCGAGCTGACCCAGCATGGTGGAAACCCCATCTCTACTAAAAA 2341
DB 241217 GAGTTCAGGAGTTCTAGACCGAGCTGACCAACATGGTGAATCTCTACTAAAAA 241158
QY 2342 CACAAAAAATTAGCGGGCGTGTGGCACATGCTGTAAATCCAGTACTCAGGAGGCTG 2401
DB 241157 TACAAAAAATTAGCGGGTGTGGTGTGGTGTGCTGTAAATCCAGCTACTCAGGAGGCTG 241098
QY 2402 AGCC-GGAGNAATGCTTGAACCCGGAAGTGGAGGTGTTGGTGTGAGCTGAGATTGCAC 2460
DB 241097 AGGCAGGAGAAATCATTGAACCTTAGAGACAGAG---GTTACAGTGTGAGCTGAGATCGCG 241041
QY 2461 TATTGCACTCCAGCTGGGCAACAGGAGTAAAACTCCCCCCCA 2503
DB 241040 CATTGCACTCCAGCTGGGCAACAGGAGTAAAACTCCCACTCA 240998
RESULT 10
ADQ97523/c
ID ADQ97523 standard; DNA; 215974 BP.
XX
AC ADQ97523;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD09-008, SEQ ID 500.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
PP WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 500; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC
```

```
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 215974 BP; 56601 A; 42897 C; 46516 G; 65586 T; 0 U; 4374 Other;
Query Match 7.5%; Score 186.8; DB 12; Length 215974;
Best Local Similarity 83.3%; Pred. No. 1.4e-32;
Matches 250; Conservative 0; Mismatches 42; Indels 8; Gaps 3;
QY 2211 AAAAATACCCCGCCAGGACGCTGAGTCCAGCCCTGTAAATCCAGCACTTTGGG----- 2264
DB 136836 ATAAATATCTCGCCAGGCGTGTGGCTCACTCTCTGTAAATCCAGCACTTTGGGAGGCCA 136777
QY 2265 AGTGTGGTGGATCACCTGAGTCCGGAGTTCAGAGCCAGCCCTGACAGCATGTGGAACC 2324
DB 136776 AGGTGGCGGATCACCTGAGTCCGGAGTTCAGAGCCAGCCCTGACCAATGG-AGAAAC 136718
QY 2325 CCCATCTCTACTAAAAACACAAAAAATTAGCCGGCGCTGGTGGCACATGCCTGTATATCCC 2384
DB 136717 CCGGTCTCTACTAAAAACACAAAAAATTAGCCAGATGTGTAGCACGCTGTATATCCC 136658
QY 2385 AGCTACTCAGGAGGCTGAGGC--GGAGAAATTGTTGAACCCGGAAGGTGGAGGTTGTCG 2443
DB 136657 AGCTACTCAGGAGACTGAGACAGGAGAAATCACTTGAATCCGGGAGGAGGTGGAGTTGCA 136598
QY 2444 GTGAGCTGAGATTGCACATTTGCACTATTTGCACTCCAGCTGGGCAACAGAGTAAAACTCCCCCCA 2503
DB 136597 GTGAGTTGACATCACGCCATTGCACTTAGCCCTGGGCAACAGAGCGAAACTCCGCTCTCA 136538
RESULT 11
AAK85292
ID AAK85292 standard; DNA; 2680 BP.
XX
AC AAK85292;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40104.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PP 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 26-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
```

PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246611P.
PR	18-AUG-2000;	2000US-0226279P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0226868P.	PR	17-NOV-2000;	2000US-0249208P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249209P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249210P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249216P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249218P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231414P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0232081P.	PR	01-DEC-2000;	2000US-0250160P.
PR	12-SEP-2000;	2000US-0232196P.	PR	01-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232399P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232400P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251989P.
PR	21-SEP-2000;	2000US-0233223P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234274P.	PR	11-DEC-2000;	2000US-0254097P.
PR	25-SEP-2000;	2000US-0234997P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0235484P.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	27-SEP-2000;	2000US-0235836P.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327P.	XX	WPI; 2001-483426/52.	
PR	29-SEP-2000;	2000US-0236327P.	DR		
PR	29-SEP-2000;	2000US-0236367P.	XX		
PR	29-SEP-2000;	2000US-0236368P.	XX		
PR	29-SEP-2000;	2000US-0236369P.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PR	29-SEP-2000;	2000US-0236370P.	PT	useful for preventing, diagnosing and/or treating cancers and metastasis.	
PR	02-OCT-2000;	2000US-0237037P.	XX	Disclosure; SEQ ID NO 40104; 3071pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237037P.	XX		
PR	02-OCT-2000;	2000US-0237038P.	XX		
PR	02-OCT-2000;	2000US-0237039P.	CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
PR	02-OCT-2000;	2000US-0237040P.	CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic	
PR	13-OCT-2000;	2000US-0239935P.	CC	activity, and can be used in gene therapy and vaccine production. (I)	
PR	13-OCT-2000;	2000US-0239937P.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
PR	20-OCT-2000;	2000US-0240960P.	CC	treatment of diseases associated with inappropriate (I) expression. For	
PR	20-OCT-2000;	2000US-0241221P.	CC	example, they may be used to treat disorders associated with decreased	
PR	20-OCT-2000;	2000US-0241785P.	CC	expression by rectifying mutations or deletions in a patient's genome	
PR	20-OCT-2000;	2000US-0241786P.	CC	that affect the activity of (I) by expressing inactive proteins or to	
PR	20-OCT-2000;	2000US-0241787P.	CC	supplement the patients own production of (I). Additionally, (I)	
PR	20-OCT-2000;	2000US-0241808P.	CC	polynucleotides may be used to produce the secreted (I), by inserting the	
PR	20-OCT-2000;	2000US-0241809P.	CC	nucleic acids into a host cell and culturing the cell to express the	
PR	20-OCT-2000;	2000US-0241826P.	CC	protein. (I) proteins and polynucleotides may be used to prevent	
PR	08-NOV-2000;	2000US-0246474P.	CC	diagnose and treat immune/haematopoietic-related diseases, especially	
PR	08-NOV-2000;	2000US-0246475P.	CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
PR	08-NOV-2000;	2000US-0246476P.	CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
PR	08-NOV-2000;	2000US-0246477P.	CC	sequences from the present invention. AAK54942 to AAK54950 and AAK82169	
PR	08-NOV-2000;	2000US-0246478P.	CC	represent sequences used in the exemplification of the present invention	
PR	08-NOV-2000;	2000US-0246523P.	XX		
PR	08-NOV-2000;	2000US-0246524P.	SQ	Sequence 2680 BP; 723 A; 594 C; 630 G; 733 T; 0 U; 0 Other;	
PR	08-NOV-2000;	2000US-0246525P.			

Query Match

7.4%; Score 186.4; DB 4; Length 2680;

Best Local Similarity 84.6%; Pred. No. 4.3e-33;
Matches 259; Conservative 0; Mismatches 36; Indels 11; Gaps 4;

QY	2205	TTATATAAAATACCCGGCCAGGCACGGTGGCTCAGCGCTGTAATCCCGACACTTTGGG	2264
Db	1131	TTTTTAAAAAGAAAGCGGCTGGCGGGTGGCTCAGCGCTGTAATCCCGACACTTTGGG	1190
QY	2265	-----AGGTGGGTGATCATCCTGAGGTGGGAGTTGCGAGCAGCGCTGACCGAGCATGGT	2318
Db	1191	AGGCCAAGTGGGCGGATCACCTGTGTCAGGAGTTGCGAGCAGCGCTGACCAACATGG-	1249
QY	2319	GGAAACCCCATCTCTACTAAAAACACAAAAATTAGCCGGCGCTGTGGGCACATGCTGT	2378
Db	1250	AGAAACCTGCTCTCTACTAAAAACACAAAAATTAGCCGGCGCTGTGGGCATGCTGT	1309
QY	2379	AATCCAGCTACTCAGGAGGCTGAGGC-GGAGAAATTGCTTTGAACCCCGAAGGTGGAGGTT	2437
Db	1310	AATCCAGCTACTCAGGAGGCTGAGGAGGAGAAATCGTTGAACCTAGGAGGCGAGAG---	1366
QY	2438	GTTCGGGTGAGTGAGATTGCACTATTGCACTCCAGCGTGGGCAACAGGAGTAAACTCC	2497
Db	1367	GTTCGAGTGAGTGAGATTGCACTCCAGCGTGGGCAACAGGAGTAAACTCC	
QY	2498	CCCCCA 2503	
Db	1427	ATCTCA 1432	

RESULT 12			
ID ABA15481/c			
AC	ABA15481;	Human nervous system related polynucleotide SEQ ID NO 7812.	
DT	23-JAN-2002	(first entry)	
DE	Human nervous system related polynucleotide SEQ ID NO 7812.		
KW	Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;		
KW	immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;		
KW	antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;		
KW	antirheumatic; hepatotropic; cerebroprotective; antinflammatory;		
KW	antiallergic; antidiabetic; antileucor; anticonvulsant; antifungal;		
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;		
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.		
OS	Homo sapiens.		
PN	WO200159063-A2.		
PD	16-AUG-2001.		
XX	17-JAN-2001; 2001WO-US001334.		
XX	31-JAN-2000; 2000US-0179065P.		
PR	04-FEB-2000; 2000US-0180628P.		
PR	24-FEB-2000; 2000US-0184664P.		
PR	02-MAR-2000; 2000US-0186350P.		
PR	16-MAR-2000; 2000US-0189874P.		
PR	17-MAR-2000; 2000US-0190076P.		
PR	18-APR-2000; 2000US-0198123P.		
PR	19-MAY-2000; 2000US-0205515P.		
PR	07-JUN-2000; 2000US-0209467P.		
PR	28-JUN-2000; 2000US-0214886P.		
PR	30-JUN-2000; 2000US-0215135P.		
PR	07-JUL-2000; 2000US-0216647P.		
PR	11-JUL-2000; 2000US-0216880P.		
PR	11-JUL-2000; 2000US-0217496P.		
PR	14-JUL-2000; 2000US-0218290P.		
PR	26-JUL-2000; 2000US-0220963P.		
PR	26-JUL-2000; 2000US-0220964P.		
PR	14-AUG-2000; 2000US-0224518P.		

PR	14-AUG-2000;	2000US-0224519P;	
PR	14-AUG-2000;	2000US-0225213P;	
PR	14-AUG-2000;	2000US-0225214P;	
PR	14-AUG-2000;	2000US-0225266P;	
PR	14-AUG-2000;	2000US-0225267P;	
PR	14-AUG-2000;	2000US-0225268P;	
PR	14-AUG-2000;	2000US-0225270P;	
PR	14-AUG-2000;	2000US-0225447P;	
PR	14-AUG-2000;	2000US-0225757P;	
PR	14-AUG-2000;	2000US-0225758P;	
PR	14-AUG-2000;	2000US-0225759P;	
PR	18-AUG-2000;	2000US-0226279P;	
PR	22-AUG-2000;	2000US-0226681P;	
PR	22-AUG-2000;	2000US-0226688P;	
PR	22-AUG-2000;	2000US-0227182P;	
PR	23-AUG-2000;	2000US-0227009P;	
PR	30-AUG-2000;	2000US-0228924P;	
PR	01-SEP-2000;	2000US-0229287P;	
PR	01-SEP-2000;	2000US-0229343P;	
PR	01-SEP-2000;	2000US-0229344P;	
PR	01-SEP-2000;	2000US-0229345P;	
PR	05-SEP-2000;	2000US-0229509P;	
PR	05-SEP-2000;	2000US-0229513P;	
PR	06-SEP-2000;	2000US-0230437P;	
PR	06-SEP-2000;	2000US-0230438P;	
PR	08-SEP-2000;	2000US-0231242P;	
PR	08-SEP-2000;	2000US-0231243P;	
PR	08-SEP-2000;	2000US-0231244P;	
PR	08-SEP-2000;	2000US-0231413P;	
PR	08-SEP-2000;	2000US-0231414P;	
PR	08-SEP-2000;	2000US-0232080P;	
PR	08-SEP-2000;	2000US-0232081P;	
PR	12-SEP-2000;	2000US-0231968P;	
PR	14-SEP-2000;	2000US-0232397P;	
PR	14-SEP-2000;	2000US-0232398P;	
PR	14-SEP-2000;	2000US-0232399P;	
PR	14-SEP-2000;	2000US-0232400P;	
PR	14-SEP-2000;	2000US-0232401P;	
PR	14-SEP-2000;	2000US-0233063P;	
PR	14-SEP-2000;	2000US-0233064P;	
PR	14-SEP-2000;	2000US-0233065P;	
PR	21-SEP-2000;	2000US-0234223P;	
PR	21-SEP-2000;	2000US-0234274P;	
PR	25-SEP-2000;	2000US-0234997P;	
PR	25-SEP-2000;	2000US-0234998P;	
PR	26-SEP-2000;	2000US-0235484P;	
PR	27-SEP-2000;	2000US-0235834P;	
PR	27-SEP-2000;	2000US-0235836P;	
PR	29-SEP-2000;	2000US-0236327P;	
PR	29-SEP-2000;	2000US-0236367P;	
PR	29-SEP-2000;	2000US-0236368P;	
PR	29-SEP-2000;	2000US-0236369P;	
PR	29-SEP-2000;	2000US-0236370P;	
PR	02-OCT-2000;	2000US-0236802P;	
PR	02-OCT-2000;	2000US-0237037P;	
PR	02-OCT-2000;	2000US-0237038P;	
PR	02-OCT-2000;	2000US-0237039P;	
PR	02-OCT-2000;	2000US-0237040P;	
PR	13-OCT-2000;	2000US-0239935P;	
PR	13-OCT-2000;	2000US-0239937P;	
PR	20-OCT-2000;	2000US-0240960P;	
PR	20-OCT-2000;	2000US-0241785P;	
PR	20-OCT-2000;	2000US-0241786P;	
PR	20-OCT-2000;	2000US-0241787P;	
PR	20-OCT-2000;	2000US-0241808P;	
PR	20-OCT-2000;	2000US-0241809P;	
PR	20-OCT-2000;	2000US-0241826P;	
PR	20-OCT-2000;	2000US-0242221P;	
PR	01-NOV-2000;	2000US-0244617P;	
PR	08-NOV-2000;	2000US-0246474P;	
PR	08-NOV-2000;	2000US-0246475P;	
PR	08-NOV-2000;	2000US-0246476P;	
PR	08-NOV-2000;	2000US-0246477P;	


```
Best Local Similarity 84.6%; Pred. No. 7.4e-33;
Matches 259; Conservative 0; Mismatches 36; Indels 11; Gaps 4;

QY 2205 TTATATAAAATACCCCGCCAGGCACGGTGGCTCAGCCTGTAAATCCAGCACTTTGGG 2264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
11708 TTTTAAAAAGAAAAGCGGCTGGGCGGGTGGCTCAGCCTGTAAATCCAGCACTTTGGG 11767

QY 2265 -----AGGTGGTGGATCACCTGAGGTGGGAGTTGAGACCAGCCTGACCAGCATGGT 2318
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
11768 AGGCCAAGGTGGGCGGATCACCTGTATGTCAGGAGTTGAGACCAGCCTGACCAACATGG- 11826

QY 2319 GGAAACCCCATCTCTACTAAAAACAAAAAATTTAGCCGGCGTGGTGGCACATGCTGT 2378
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
11827 AGAAACCTCGTCTCTACTAAAAACAAAAAATTTAGCCGGCGTGGTGGCCATGCTGT 11886

QY 2379 AATCCAGCTACTCAGGAGCTGAGG- GGAGAAATTGCTTGAACCCGGAAGGTGGAGGTT 2437
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
11887 AATCCAGCTACTCAGGAGCTGAGGAGGAGAAATCGTTGAACTTAGGAGGAGAG--- 11943

QY 2438 GTTGGGTGAGCTGAGATTGCACTATTGCATCCAGCCTGGGCAACAGGATAAACTCC 2497
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
11944 GTTGCAGTGAGCTGAGATTGCACCTGCACCCAGCCTGGGCAATAGAGTGAATACTCC 12003

QY 2498 CCCCCA 2503
Db |||||
12004 ATCTCA 12009
```

Search completed: September 8, 2005, 23:49:50
Job time : 894.428 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 23:15:40 ; Search time 289.657 Seconds
(without alignments)
14156.413 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gacctaataatagaggt.....agtaaaactccccccacc 2506

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCUTS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2494	99.5	2505	3	US-09-097-199-85 Sequence 85, Appl
2	2424.2	96.7	2506	4	US-09-949-016-1311 Sequence 3311, Ap
3	2424.2	96.7	6507	4	US-09-949-016-15053 Sequence 15053, A
4	1950	77.8	2087	3	US-09-097-199-83 Sequence 83, Appl
5	626.2	25.0	757	2	US-08-692-787-3 Sequence 3, Appl1
6	572.2	22.8	601	4	US-09-949-016-119365 Sequence 119365, A
7	572.2	22.8	601	4	US-09-949-016-119366 Sequence 119366, A
8	572.2	22.8	601	4	US-09-949-016-119367 Sequence 119367, A
9	397.6	15.9	601	4	US-09-949-016-119364 Sequence 119364, A
10	344.2	13.7	601	4	US-09-949-016-119364 Sequence 119364, A
11	189.4	7.6	174639	4	US-09-949-016-16509 Sequence 121999, A
12	189	7.5	601	4	US-09-949-016-16509 Sequence 16509, A
13	188	7.5	48691	4	US-09-949-016-16308 Sequence 16308, A
14	187.8	7.5	70947	4	US-09-949-016-15165 Sequence 15165, A
15	187.6	7.5	601	4	US-09-949-016-163069 Sequence 163069, A
16	187.6	7.5	601	4	US-09-949-016-163070 Sequence 163070, A
17	187.4	7.5	601	4	US-09-949-016-122000 Sequence 122000, A
18	186.6	7.4	27630	4	US-09-949-016-12722 Sequence 12722, A
19	185.4	7.4	571	4	US-09-621-976-17222 Sequence 17222, A
20	185.4	7.4	209210	4	US-09-949-016-15094 Sequence 15094, A
21	185	7.4	601	4	US-09-949-016-121998 Sequence 121998, A
22	185	7.4	601	4	US-09-949-016-163071 Sequence 163071, A
23	184.8	7.4	69062	4	US-09-949-016-13608 Sequence 13608, A
24	184.8	7.4	69062	4	US-09-949-016-13609 Sequence 13609, A
25	184.8	7.4	69687	4	US-09-949-016-12890 Sequence 12890, A
26	184.4	7.4	601	4	US-09-949-016-63806 Sequence 63806, A
27	184.4	7.4	601	4	US-09-949-016-63881 Sequence 63881, A

C	28	184.2	7.4	565	4	US-09-621-976-17221 Sequence 17221, A
	29	183.6	7.3	106746	4	US-09-326-402C-1 Sequence 1, Appl
	30	183.6	7.3	106746	4	US-09-326-402C-12 Sequence 12, Appl
	31	183.4	7.3	75212	4	US-09-949-016-13313 Sequence 13313, A
	32	183.4	7.3	75212	4	US-09-949-016-13314 Sequence 13314, A
	33	183.4	7.3	75212	4	US-09-949-016-13315 Sequence 13315, A
	34	183.4	7.3	75212	4	US-09-949-016-13670 Sequence 12670, A
	35	183.4	7.3	87523	4	US-09-949-016-15047 Sequence 15047, A
	36	183.4	7.3	87523	4	US-09-949-016-15048 Sequence 15048, A
	37	183.4	7.3	87523	4	US-09-949-016-15049 Sequence 15049, A
	38	183.4	7.3	87869	4	US-09-949-016-11744 Sequence 11744, A
	39	183.4	7.3	87869	4	US-09-949-016-15044 Sequence 15044, A
	40	183.4	7.3	87869	4	US-09-949-016-15045 Sequence 15045, A
	41	183.4	7.3	87869	4	US-09-949-016-15046 Sequence 15046, A
C	42	183.2	7.3	19728	4	US-09-949-016-12506 Sequence 12506, A
C	43	183.2	7.3	19728	4	US-09-949-016-15767 Sequence 15767, A
C	44	182.6	7.3	51671	4	US-09-949-016-12068 Sequence 12068, A
C	45	182.6	7.3	51671	4	US-09-949-016-15962 Sequence 15962, A

ALIGNMENTS

RESULT 1
US-09-097-199-85
; Sequence 85, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2505 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 99..503
US-09-097-199-85

Query Match 99.5%; Score 2494; DB 3; Length 2505;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2505; Conservative 0; Mismatches 0; Indels 1; Gaps 1;			
QY	1	GACCTTAAATATATCGAGGTGGCTAATTGATGTATTAATAATTACAAAATATTCTTCTTA	60
Db	1	GACCTTAAATATATCGAGGTGGCTAATTGATGTATTAATAATTACAAAATATTCTTCTTA	60
QY	61	TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCATGAGGGCTTCTTAAAGAAC	120
Db	61	TTGCTACAGAGCTACAAATTCAAATTTACAGTAGGCCACCATGAGGGCTTCTTAAAGAAC	120
QY	121	AGAAATATGAGGATATGCAAAATATTATTACAAATTTTACAGATCAGAAAATTTGAGSCACA	180
Db	121	AGAAATATGAGGATATGCAAAATATTATTACAAATTTTACAGATCAGAAAATTTGAGSCACA	180
QY	181	GATTAAAGTAACTTCCCAAGGCTACAGGCATTCAGCTCCAGNAACTGTGCTTTACCAT	240
Db	181	GATTAAAGTAACTTCCCAAGGCTACAGGCATTCAGCTCCAGNAACTGTGCTTTACCAT	240
QY	241	TCTGCTACAAAGGTATTTCGAAAAAAGAAAAAGTAAAAAGAGTCAAAAAGGCCAACAGAGT	300
Db	241	TCTGCTACAAAGGTATTTCGAAAAAAGAAAAAGTAAAAAGAGTCAAAAAGGCCAACAGAGT	300
QY	301	TCATTGATTAATCCATAGAACAGTCAACATGCAATTTCTCACCCCTTGCAGACACACT	360
Db	301	TCATTGATTAATCCATAGAACAGTCAACATGCAATTTCTCACCCCTTGCAGACACACT	360
QY	361	TGACCATTGAAGGTTCCCTCAATGAAATGTTCCCTCATTATCTTCAGAGGCCATATTATCCA	420
Db	361	TGACCATTGAAGGTTCCCTCAATGAAATGTTCCCTCATTATCTTCAGAGGCCATATTATCCA	420
QY	421	CATTGACTTTGCGAGTTAACTCAGACCCCTAGGTCGGAATGCTGTCTCTCTACTTATCCA	480
Db	421	CATTGACTTTGCGAGTTAACTCAGACCCCTAGGTCGGAATGCTGTCTCTCTACTTATCCA	480
QY	481	AAACTATATACATCCAGATCATATAAATCTCTCAGCCCTGCTGCAAGCCCTTCCAGAAAA	540
Db	481	AAACTATATACATCCAGATCATATAAATCTCTCAGCCCTGCTGCAAGCCCTTCCAGAAAA	540
QY	541	ATAAAAATGGTTGAAAAGGCAATTCGCTACCAATGACTGTTTAAGCCCGAGCCAGTAAAC	600
Db	541	ATAAAAATGGTTGAAAAGGCAATTCGCTACCAATGACTGTTTAAGCCCGAGCCAGTAAAC	600
QY	601	TGAACCAATTCCTCAATTTACTTATCAAAAAGAAATTTGATGATGAGAGGTTATTTTC	660
Db	601	TGAACCAATTCCTCAATTTACTTATCAAAAAGAAATTTGATGATGAGAGGTTATTTTC	660
QY	661	AATTTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTAATTATCT	720
Db	661	AATTTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAACTCATAGATTAATTATCT	720
QY	721	ATTATCTCAATTTAGTTTCTGTTATTTATCTAGTGGGCCATTAARAACTACCACATGTGTT	780
Db	721	ATTATCTCAATTTAGTTTCTGTTATTTATCTAGTGGGCCATTAARAACTACCACATGTGTT	780
QY	781	TCTGCTCTCCATTAAGTCAATAAATTAACGAGCAATTTAGTAGCCATGTGCCAGAT	840
Db	781	TCTGCTCTCCATTAAGTCAATAAATTAACGAGCAATTTAGTAGCCATGTGCCAGAT	840
QY	841	GCTCCGCTAGGCCACGAGGGATATAAAAACAATCTTATAGTATACCACTAAATTTTCGCTT	900
Db	841	GCTCCGCTAGGCCACGAGGGATATAAAAACAATCTTATAGTATACCACTAAATTTTCGCTT	900
QY	901	AGTAACTAGTGAATTTCTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTTACAACTGT	960
Db	901	AGTAACTAGTGAATTTCTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTTACAACTGT	960
QY	961	GTAATGAAAACCAAGGAAAGTGAAAATTTTGGATAAGTGGGGACTAGTGTATTATATTT	1020
Db	961	GTAATGAAAACCAAGGAAAGTGAAAATTTTGGATAAGTGGGGACTAGTGTATTATATTT	1020
QY	1021	TAATTGATTTCTGACTCTATCATTTGGCCCTCAGAAACACAGATTTGTGTTTTCTTTGGTTTT	1080

Db	1021	TAATTGATTTCTGACTCTATCATTTGGCCCTCAAAACACAGATTTGTGTTTTCTTTGGTTTT	1080
QY	1081	GTTTTCTTTCACATATGGGATCTTCTGTGCCAGCACAGTGCCTGACACATAGAAAACAATC	1140
Db	1081	GTTTTCTTTCACATATGGGATCTTCTGTGCCAGCACAGTGCCTGACACATAGAAAACAATC	1140
QY	1141	AATATTTGCTGTAATAAATGATTTAAAAAATCAGAGAACTTTCCCATCTCTGTTTGGATCTAT	1200
Db	1141	AATATTTGCTGTAATAAATGATTTAAAAAATCAGAGAACTTTCCCATCTCTGTTTGGATCTAT	1200
QY	1201	AGAACTCCAGAGTAAGTATGAGGGCTCTGCAATTTATATGCGCTTAAATTAAGATTAAT	1260
Db	1201	AGAACTCCAGAGTAAGTATGAGGGCTCTGCAATTTATATGCGCTTAAATTAAGATTAAT	1260
QY	1261	GTGAGAAAGTTTAAAGACACTTAGTAGAGTGAATTTGAAATATATAGTAAACACTTTGGAAA	1320
Db	1261	GTGAGAAAGTTTAAAGACACTTAGTAGAGTGAATTTGAAATATATAGTAAACACTTTGGAAA	1320
QY	1321	TGGTGGTCTTTTAAAGAGATATTAATAGATAATATGAAAAATCTCCATCTCAAAAAATAATG	1380
Db	1321	TGGTGGTCTTTTAAAGAGATATTAATAGATAATATGAAAAATCTCCATCTCAAAAAATAATG	1380
QY	1381	CATAAATCTATTTAAAGGAAAATCAATCTCAGGGCTTTCAATGTTTGTTCATTCTTTT	1440
Db	1381	CATAAATCTATTTAAAGGAAAATCAATCTCAGGGCTTTCAATGTTTGTTCATTCTTTT	1440
QY	1441	CATATATTTTACCATCTGCTGAGGCGAGTCAATCAAGGGTAAAGAAAGATGGAGGA	1500
Db	1441	CATATATTTTACCATCTGCTGAGGCGAGTCAATCAAGGGTAAAGAAAGATGGAGGA	1500
QY	1501	AAACTCAGTAAGAAATATATTAATGATCTGTTTGCAAAAGTAGAAAAAGATTTCTCATCTCAA	1560
Db	1501	AAACTCAGTAAGAAATATATTAATGATCTGTTTGCAAAAGTAGAAAAAGATTTCTCATCTCAA	1560
QY	1561	CTTATAGCAGGAAAGAGGAGGCTGTTTGAGAAACATTTTCTTAGCAGAAACCAATAT	1620
Db	1561	CTTATAGCAGGAAAGAGGAGGCTGTTTGAGAAACATTTTCTTAGCAGAAACCAATAT	1620
QY	1621	TTTACACACTTCCCTGCAATTAACATCAACAAATATGTTTGCRAACTTGTTRGATCAAC	1680
Db	1621	TTTACACACTTCCCTGCAATTAACATCAACAAATATGTTTGCRAACTTGTTRGATCAAC	1680
QY	1681	CTCCAAACACACACATTTGAGGAGTAAATATTTTTCATCAAAACATTTGGATTTTTCCTTA	1740
Db	1680	CTCCAAACACACACATTTGAGGAGTAAATATTTTTCATCAAAACATTTGGATTTTTCCTTA	1739
QY	1741	ACGCTAGAGATTGCTACAAATCTTCTGAAAGGCTCTCAATGGGCTTCAGGCTTAAGAGAGAT	1800
Db	1740	ACGCTAGAGATTGCTACAAATCTTCTGAAAGGCTCTCAATGGGCTTCAGGCTTAAGAGAGAT	1799
QY	1801	TTCTCCCTGTTTATAGCAGCAAGCAAAATTAGCCATTTTCACTCTCAAACTTCACCTAATGA	1860
Db	1800	TTCTCCCTGTTTATAGCAGCAAGCAAAATTAGCCATTTTCACTCTCAAACTTCACCTAATGA	1859
QY	1861	TCACATTTCTTTTCAAAAGGAACTCTAGAGACCCAAATGCCCGAGTTTAAAGAACATCAAAA	1920
Db	1860	TCACATTTCTTTTCAAAAGGAACTCTAGAGACCCAAATGCCCGAGTTTAAAGAACATCAAAA	1919
QY	1921	CTAACCTCTGAGAAACTTCCCAAGTGAAGACTCTGCCATTAATAAATTTACCCGAGAGG	1980
Db	1920	CTAACCTCTGAGAAACTTCCCAAGTGAAGACTCTGCCATTAATAAATTTACCCGAGAGG	1979
QY	1981	GGACTCAAAACAGTCTTTCTCTTTGCTGTTTCTGCTCCAGACCAAGGCACTGAGG	2040
Db	1980	GGACTCAAAACAGTCTTTCTCTTTGCTGTTTCTGCTCCAGACCAAGGCACTGAGG	2039
QY	2041	ACAGTACTGATACATAAATTTTAAAGCACACTCCCTTCCACTTTGGTAAATACAGAACTCT	2100
Db	2040	ACAGTACTGATACATAAATTTTAAAGCACACTCCCTTCCACTTTGGTAAATACAGAACTCT	2099
QY	2101	AAATGGACCAACCTGAGCTTAGGACTACAGGCCATACAAATAGTAAACTCTGTCCAGGA	2160
Db	2100	AAATGGACCAACCTGAGCTTAGGACTACAGGCCATACAAATAGTAAACTCTGTCCAGGA	2159

```
QY 2161 TTCACTCATCTGTGATTTTCTATAGATGTTTACTAGGCGTTGTTATATAAAAAATACCC 2220
DB 2160 TTCACTCATCTGTGATTTTCTATAGATGTTTACTAGGCGTTGTTATATAAAAAATACCC 2219
QY 2221 CGGCCAGGACGGTGGCTCAACGCTGCTAAATCCAGCACATTTGGGAGGTGGGTGGATCACC 2280
DB 2220 CGGCCAGGACGGTGGCTCAACGCTGCTAAATCCAGCACATTTGGGAGGTGGGTGGATCACC 2279
QY 2281 TGAGGTGGGAGTTTCAGACACGCTGACAGCATGGTGGAAACCCCACTCTCTACTAAAA 2340
DB 2280 TGAGGTGGGAGTTTCAGACACGCTGACAGCATGGTGGAAACCCCACTCTCTACTAAAA 2339
QY 2341 ACACAAAAATTAAGCCGGCGGTGGTGCACATCCCTGTAATCCAGCTACTCAGGAGGCT 2400
DB 2340 ACACAAAAATTAAGCCGGCGGTGGTGCACATCCCTGTAATCCAGCTACTCAGGAGGCT 2399
QY 2401 GAGCGCGAGAAATGCTTGAACCCGGAGGTGGAGGTGTTGGCGTGCAGCTGAGATTGCAC 2460
DB 2400 GAGCGCGAGAAATGCTTGAACCCGGAGGTGGAGGTGTTGGCGTGCAGCTGAGATTGCAC 2459
QY 2461 TATTGCACCTCCAGCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2506
DB 2460 TATTGCACCTCCAGCTGGGCAACAGGAGTAAAACTCCCCCCCACCC 2505

RESULT 2
US-09-949-016-3311
; Sequence 3311, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3311
; LENGTH: 2506
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-3311

Query Match 96.7%; Score 2424.2; DB 4; Length 2506;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2494; Conservative 1; Mismatches 4; Indels 8; Gaps 6;

QY 1 GACCTTAATATATCGAGGTGGCTAATTGATGATATAATTTACAAAATATCTTCTA 60
DB 1 GACCTTAATATATCGAGGTGGCTAATTGATGATATAATTTACAAAATATCTTCTA 60
QY 61 TTGCTACAGAGCTACAAATTTCAATTTACAGTACGACCCCACTAGGCGCTTCTTAAGAAC 120
DB 61 -TGCTACAGAGCTACAAATTTCAATTTACAGTA -GTAATATGAGGCGCTTCTTAAGAAC 118
QY 121 AGAAATATGAGGATATGCACAATATTTATTCACATTTTACAGATCAGAAAATTCAGGCACA 180
DB 119 AGAAATATGAGGATATGCACAATATTTATTCACATTTTACAGATCAGAAAATTCAGGCACA 178
QY 181 GATTAGTAATCTCCAGGCTACAGGCAATCTAGCTCCAGAAATCTGTGCTCTTACCAT 240
DB 179 GATTAGTAATCTCCAGGCTACAGGCAATCTAGCTCCAGAAATCTGTGCTCTTACCAT 238
QY 241 TCTGCTACAGGTATTTTCGAAAAAGCAAGTAAAAAGTCAAAAGGCAACAGCT 300
```

```
DB 239 TCTGCTACAGGTATTTTCGAAAAAGCAAGTAAAAAGAGTCAAAGGCAACAGAGT 298
QY 301 TCATGATTAATTCATAGAAACAGTCAACCATGCAATTTCTCAACCTTGCAGACACACT 360
DB 299 TCATGATTAATTCATAGAAACAGTCAACCATGCAATTTCTCAACCTTGCAGACACACT 358
QY 361 TCAGCATGAAGGTTCCCTCAATGAATGTTCCCTCAATTAATCTTCAGAAAGCATATATTCA 420
DB 359 TCAGCATGAAGGTTCCCTCAATGAATGTTCCCTCAATTAATCTTCAGAAAGCATATATTCA 418
QY 421 CATTGACTTTGCAGTTAACTCAGACCCCTAGGCTCTGGAATGCTGTCTTCTCTACTTATCA 480
DB 419 CATTGACTTTGCAGTTAACTCAGACCCCTAGGCTCTGGAATGCTGTCTTCTCTACTTATCA 478
QY 481 AAATATATACATCCAGATCATATAAACTCTCAGCCCTGCTGCAAGCCCTTTCAGAAAA 540
DB 479 AAATATATACATCCAGATCATATAAACTCTCAGCCCTGCTGCAAGCCCTTTCAGAAAA 538
QY 541 ATAAAAATGTTTGAAGGCAATCTGCTACCAATGACTGTTTAAAGCCAGCCCAAGTAAC 600
DB 539 ATAAAAATGTTTGAAGGCAATCTGCTACCAATGACTGTTTAAAGCCAGCCCAAGTAAC 598
QY 601 TGAACCATTTCCAACTTCAATTTTACTTATGAAAGAAATTTGATGATGTAGAGGTTATTTTC 660
DB 599 TGAACCATTTCCAACTTCAATTTTACTTATGAAAGAAATTTGATGATGTAGAGGTTATTTTC 658
QY 661 AATTTCTAAAAATACAAACCCATGTTGATCTTCTCAATCTTGAACCTCATAGATATTATCT 720
DB 659 AATTTCTAAAAATACAAACCCATGTTGATCTTCTCAATCTTGAACCTCATAGATATTATCT 718
QY 721 ATTATCTCAATTTAGTTTGTATTTATCTCTAGTGGGCCATTAAGCAATAGTAAACCTACATGTT 780
DB 719 ATTATCTCAATTTAGTTTGTATTTATCTCTAGTGGGCCATTAAGCAATAGTAAACCTACATGTT 778
QY 781 TCTGCTCTCTCCATTAAGTCAATAAATAAAGCAATTAAGTAAAGCCATGTGCCAGAT 840
DB 779 TCTGCTCTCTCCATTAAGTCAATAAATAAAGCAATTAAGTAAAGCCATGTGCCAGAT 838
QY 841 GTCGCTAGGACACAGAGGGATAAAAACAATATCTATAGTATACACATAATTTTCGCTT 900
DB 839 GTCGCTAGGACACAGAGGGATAAAAACAATATCTATAGTATACACATAATTTTCGCTT 898
QY 901 AGTAACTAGTGAATGTTCAAGTCAATGCTGAGTCAAGGTTGAGGAGACATTTACAATCT 960
DB 899 AGTAACTAGTGAATGTTCAAGTCAATGCTGAGTCAAGGTTGAGGAGACATTTACAATCT 958
QY 961 GTAATGGAAACCAAGAAAGTCAAACTTTGGATAAGTGGGACTAGTGTATTATATATT 1020
DB 959 GTAATGGAAACCAAGAAAGTCAAACTTTGGATAAGTGGGACTAGTGTATTATATATT 1018
QY 1021 TAATGATTTCTGACTCTATCAATTTGGCTCCAAAACAGATTTGTTTCTTTGGTTTT 1080
DB 1019 TAATGATTTCTGACTCTATCAATTTGGCTCCAAAACAGATTTGTTTCTTTGGTTTT 1078
QY 1081 GTTTTCTTCACTATGGGATCTTCTGCCCCAGACAGTCCCTGACACATAGAAAACAATC 1140
DB 1079 GTTTTCTTCACTATGGGATCTTCTGCCCCAGACAGTCCCTGACACATAGAAAACAATC 1138
QY 1141 -AATATTTGCTCAATAAATGATTAATAAATAACAGAGAACTTTCCCAATTTCTGTTTGAATCTA 1199
DB 1139 AATATTTGCTCAATAAATGATTAATAAATAACAGAGAACTTTCCCAATTTCTGTTTGAATCTA 1198
QY 1200 TAGAAATCCAGAGTAAGTATGAGGCGCTCTGCAATTTATATGCGCTTAAATTAAGATTA 1259
DB 1199 TAGAAATCCAGAGTAAGTATGAGGCGCTCTGCAATTTATATGCGCTTAAATTAAGATTA 1258
QY 1260 TGTGAGAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAATATAGTAAACACTTGGAA 1319
DB 1259 TGTGAGAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAATATAGTAAACACTTGGAA 1318
QY 1320 ATGTGGTGTCTTTTAAAGAGATTAATATAGATTAATGAAAACTCTCCATCTCAAAAATAAT 1379
DB 1319 ATGTGGTGTCTTTTAAAGAGATTAATATAGATTAATGAAAACTCTCCATCTCAAAAATAAT 1378
```


QY 601 TGAACCATTCACACTTCAATTTACTTATGAAAGAAATTTGATGATGTAGGAGTTATTTTC 660
DB 2599 TGAACCATTCACACTTCAATTTACTTATGAAAGAAATTTGATGATGTAGGAGTTATTTTC 2658
QY 661 AATTTCTAAATAACAAACCCAGTTGTATCTTTCTCAATCTTGAACCTCATAGATTTATTTATCT 720
DB 2659 AATTTCTAAATAACAAACCCAGTTGTATCTTTCTCAATCTTGAACCTCATAGATTTATTTATCT 2718
QY 721 ATTATCTCAATTTAGTTTGTATTTATTTCTAGTGGGCCATTTAAAACTACCAATGTGTT 780
DB 2719 ATTATCTCAATTTAGTTTGTATTTATTTCTAGTGGGCCATTTAAAACTACCAATGTGTT 2778
QY 781 TCTGTCTCTCCATTTAGTCAATTAATACTAACTAAAGAGCAATTTAGTAAAGCCATGTGCCAGAT 840
DB 2779 TCTGTCTCTCCATTTAGTCAATTAATACTAACTAAAGAGCAATTTAGTAAAGCCATGTGCCAGAT 2838
QY 841 GCTCCGCTAGGCACACAGAGGATATAAAACAATTAATTTATAGTATACCACTAAATTTTCGCTT 900
DB 2839 GCTCCGCTAGGCACACAGAGGATATAAAACAATTAATTTATAGTATACCACTAAATTTTCGCTT 2898
QY 901 AGTAACTAGTGAATTTTCAAGTCTATGCTGCTCAAGAGTTGAGGAGACATTTACAAATGT 960
DB 2899 AGTAACTAGTGAATTTTCAAGTCTATGCTGCTCAAGAGTTGAGGAGACATTTACAAATGT 2958
QY 961 GTAATGGAACCAAGAAAGTGAACCTTTTGGATAAGTGGGACTAGTGTATTTATATATTT 1020
DB 2959 GTAATGGAACCAAGAAAGTGAACCTTTTGGATAAGTGGGACTAGTGTATTTATATATTT 3018
QY 1021 TAATTCATTTCTGACTCTATCATTTGGCTCCAAACACAGATTTGTTTCTTTGGTTTTT 1080
DB 3019 TAATTCATTTCTGACTCTATCATTTGGCTCCAAACACAGATTTGTTTCTTTGGTTTTT 3078
QY 1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGACACAGTGCCTGACATAGAAAAACAATC 1140
DB 3079 GTTTTCTTCACTATGGGATCTTCTGTGCCAGACACAGTGCCTGACATAGAAAAACAATC 3138
QY 1141 -AATATTTGCTGAATTAATGAATTAATAAATACAGAGAACTTTTCCCATCTGTTTGGATCTA 1199
DB 3139 AATATTTTGTGTAATTAATGAATTAATAAATACAGAGAACTTTTCCCATCTGTTTGGATCTA 3198
QY 1200 TAGAACATCCAGAGTAAGTATGAGGCTCTGCAATTTATATGCGCTTAAATTTAAGATTA 1259
DB 3199 TAGAACATCCAGAGTAAGTATGAGGCTCTGCAATTTATATGCGCTTAAATTTAAGATTA 3258
QY 1260 TGTGAGAAAGTTTAAAGACACTTTAGTAGAGTGAATTTGAAATATATAGTAAACACTTTGGAA 1319
DB 3259 TGTGAGAAAGTTTAAAGACACTTTAGTAGAGTGAATTTGAAATATATAGTAAACACTTTGGAA 3318
QY 1320 ATGGTGGTCTTTAAAGAGATATTAATAGATATATGAAATCTCCATCTCAAAAAATAAT 1379
DB 3319 ATGGTGGTCTTTAAAGAGATATTAATAGATATATGAAATCTCCATCTCAAAAAATAAT 3378
QY 1380 GCATAAATCTTTAAAGGAAATACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 1439
DB 3379 GCATAAATCTTTAAAGGAAATACATCTCCAGGCTTTCAATGTTTGTTCATTACTTTT 3438
QY 1440 TCATATATTTTACCATCTGCTGAAGGCAGTCAATATCAAGGGTAAAGAGATGGGAGG 1499
DB 3439 TCATATATTTTACCATCTGCTGAAGGCAGTCAATATCAAGGGTAAAGAGATGGGAGG 3498
QY 1500 AAACTCAGTAAGATTAATTAATAGTCTGTTTCAAAAGTGAAGAAAGTTCTCATCACTCA 1559
DB 3499 AAACTCAGTAAGATTAATTAATAGTCTGTTTCAAAAGTGAAGAAAGTTCTCATCACTCA 3558
QY 1560 ACCATTAGCAGGAGAGGGAGGCTTTTGAAGACCATTTACTTAGCAGAACCATATA 1619
DB 3559 ACCATTAGCAGGAGAGGGAGGCTTTTGAAGACCATTTACTTAGCAGAACCATATA 3618
QY 1620 TTTTAGACACTTCCCTGCAATTAATCTGACAAAAATAATGTTTCAAACTGTTTGTATCAA 1679
DB 3619 TTTTAGACACTTCCCTGCAATTAATCTGACAAAAATAATGTTTCAAACTGTTTGTATCAA 3678

QY 1680 CCTCCAAACAAACGACACATTCAGGAGTTAAATATTTTTCATCAAAATTTGGATTTTTCCTT 1739
DB 3679 CCTCCAAACAAACGACACATTCAGGAGTTAAATATTTTTCATCAAAATTTGGATTTTTCCTT 3738
QY 1740 AACGCTAGAGATTTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAAGAGA 1799
DB 3739 AACGCTAGAGATTTGCTACAAATCTTCTGAAGGCTCTCAATGGCTTCAGGCTAAGAAGAGA 3798
QY 1800 TTTTCTCCCTGTTTATAGCAGCAAGACAAATTAAGCCATTTCTCTCAAACTTCACTAATG 1859
DB 3799 TTTTCTCCCTGTTTATAGCAGCAAGACAAATTAAGCCATTTCTCTCAAACTTCACTAATG 3858
QY 1860 ATCAATCTTTTCCAAAGGAACTCTAGAGACCAATGCCCCGGAGTTAAGAACATCAAA 1919
DB 3859 ATCAATCTTTTCCAAAGGAACTCTAGAGACCAATGCCCCGGAGTTAAGAACATCAAA 3918
QY 1920 ACTAACCATCTCAAGAACTTCCCAAGTGTAGAGACTCTGCCATTAATAAATTTACCGAGAG 1979
DB 3919 ACTAACCATCTCAAGAACTTCCCAAGTGTAGAGACTCTGCCATTAATAAATTTACCGAGAG 3978
QY 1980 GGGACTCAAAACAGTCTTTCTTCTTGTGCTGTTTCTTGTCTCCAGACCAAGGCACTGAC 2039
DB 3979 GGGACTCAAAACAGTCTTTCTTCTTGTGCTGTTTCTTGTCTCCAGACCAAGGCACTGAC 4038
QY 2040 GACAGTACTGATACATATTTTAAAGCACACTCCCTTCCACTTTTGTATATACCAACTC 2099
DB 4039 GACAGTACTGATACATATTTTAAAGCACACTCCCTTCCACTTTTGTATATACCAACTC 4097
QY 2100 TAAATTTGGACACCTTGAAGCTTAGGACTACAGCAGCATCAAAATAGTAAATCTGTCCACG 2159
DB 4098 TAAATTTGGA-CACCTTGAAGCTTAGGACTACAGCAGCATCAAAATAGTAAATCTGTCCACG 4156
QY 2160 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTATATATAAATAACC 2219
DB 4157 ATTCACTCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTATATATAAATAACC 4216
QY 2220 CCGGCGGACGAGTGGCTCAGCCCTGTAAATCCAGCAGCTTTGGGAGGTGGGTGATCAC 2279
DB 4217 CCGGCGGACGAGTGGCTCAGCCCTGTAAATCCAGCAGCTTTGGGAGGTGGGTGATCAC 4276
QY 2280 CTGAGTCTGGGAGTTTCGAGACAGCAGCTGACAGCAGTGTGGAAACCCCATCTCTACTAAA 2339
DB 4277 CTGAGTCTGGGAGTTTCGAGACAGCAGCTGACAGCAGTGTGGAAACCCCATCTCTACTAAA 4336
QY 2340 AACACAAAAAATTAGCCGGGCTGTGGCAGATGCTGTAAATCCAGCTTACAGGAGGC 2399
DB 4337 AACACAAAAAATTAGCCGGGCTGTGGCAGATGCTGTAAATCCAGCTTACAGGAGGC 4396
QY 2400 TGAGGGGAGGATTTCTTGAACCCGGAGGTGGAGTTGTTGGGTGAGCTGAGATTGCA 2459
DB 4397 TGAGGGGAGGATTTCTTGAACCCGGAGGTGGAG---GTTGGGTGAGCTGAGATTGCA 4453
QY 2460 CTATTGCACTCCAGCTGGGCAACAGGAGTAAATCTCCCCCCCCACCC 2506
DB 4454 CTATTGCACTCCAGCTGGGCAACAGGAGTAAATCTCCCCCCCCACCC 4500

RESULT 4

US-09-199-83
; Sequence 83, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltre, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

```
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/097,199
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/692,787
/ FILING DATE: 31-JUL-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 83:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2087 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 99..503
/ US-09-097-199-83

Query Match 77.8%; Score 1950; DB 3; Length 2087;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1967; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 GACCTTAAATATATCGAGGTGGCTAATTCATGTATATAATTAATTAACAAAATTAATCTCTTA 60
DB 1 GACCTTAAATATATCGAGGTGGCTAATTCATGTATATAATTAATTAACAAAATTAATCTCTTA 60

QY 61 TTGCTACAGAGTACAAATTCAAATTTACAGTAGGCCACCATAGAGGCCCTTTTAAAGAAC 120
DB 61 TTGCTACAGAGTACAAATTCAAATTTACAGTAGGCCACCATAGAGGCCCTTTTAAAGAAC 120

QY 121 AGAAATATGAGGATATGCACATATATTTACATTTTACAGATCAGAAAATTTGAGGCACA 180
DB 121 AGAAATATGAGGATATGCACATATATTTACATTTTACAGATCAGAAAATTTGAGGCACA 180

QY 181 GATTAAAGTAACCTTCCAAAGGCTACAGGCATTTAGCTCCAGAAACTGTGCTCTTACCAT 240
DB 181 GATTAAAGTAACCTTCCAAAGGCTACAGGCATTTAGCTCCAGAAACTGTGCTCTTACCAT 240

QY 241 TCTGCTACAGGTATTTTCAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 300
DB 241 TCTGCTACAGGTATTTTCAAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT 300

QY 301 TCATTGATTTATCCATAGAACAGTACACCATGCAATTTCTCACCCCTTGCAGACACACT 360
DB 301 TCATTGATTTATCCATAGAACAGTACACCATGCAATTTCTCACCCCTTGCAGACACACT 360

QY 361 TGACCATGAAAGGTTCTCAATGAATGTTCTCTCATTTATCTTTAGAGGCATATTTATTTCA 420
DB 361 TGACCATGAAAGGTTCTCAATGAATGTTCTCTCATTTATCTTTAGAGGCATATTTATTTCA 420

QY 421 CATTTGACTTTGAGTTAACTCAGACCCCTAGGTCTGGAATGCTGCTCTCTCTATCTTATCCA 480
DB 421 CATTTGACTTTGAGTTAACTCAGACCCCTAGGTCTGGAATGCTGCTCTCTCTCTATCTTATCCA 480

QY 481 AAACATATACATCCACAGATCATATAAATCTCAGCCCTCCTCAGAGCCCTTTCCAGAAAA 540
DB 481 AAACATATACATCCACAGATCATATAAATCTCAGCCCTCCTCAGAGCCCTTTCCAGAAAA 540

541 ATAAAAATGGTTGAAAAGGCAATTCCTGCTACCAATGACTGTTTAAGCCAGCCAGTAAC 600
541 ATAAAAATGGTTGAAAAGGCAATTCCTGCTACCAATGACTGTTTAAGCCAGCCAGTAAC 600

601 TGAACCATTCGAACTTCAATTTTATGATGAAAAGAAATTTGATGATGATGATGATGATGATG 660
601 TGAACCATTCGAACTTCAATTTTATGATGAAAAGAAATTTGATGATGATGATGATGATGATG 660

661 AATCTTAAATACAAACCCATGTTGATCTTCTCAATCTTGAACCTCATAGATTATATCT 720
661 AATCTTAAATACAAACCCATGTTGATCTTCTCAATCTTGAACCTCATAGATTATATCT 720

721 ATTATCTCAATTTAGTTTGTATTTATCTTAGTGGGCCAATTAATAAATCTACCATGTGTT 780
721 ATTATCTCAATTTAGTTTGTATTTATCTTAGTGGGCCAATTAATAAATCTACCATGTGTT 780

781 TCTGCTCTCCATTTAGTCAATAAATAAAGCAATTAAGTAAAGCAATTTAGTAAAGCAATTTAG 840
781 TCTGCTCTCCATTTAGTCAATAAATAAAGCAATTTAGTAAAGCAATTTAGTAAAGCAATTTAG 840

841 GCTCCGCTAGGACCCAGAGGGATAAAAACAATACCTTATAGTATACCACTAATTTTCGCTT 900
841 GCTCCGCTAGGACCCAGAGGGATAAAAACAATACCTTATAGTATACCACTAATTTTCGCTT 900

901 AGTAACTAGTGAAATGTTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTTACAATGT 960
901 AGTAACTAGTGAAATGTTTCAAGTCAATGCTGAGTCAAGAGTTGAGGAGACATTTACAATGT 960

961 GTAATGGAACCAAGGAAAGTGAAACTTTTGATTAAGTGGGACTAGTGTATTTATATAT 1020
961 GTAATGGAACCAAGGAAAGTGAAACTTTTGATTAAGTGGGACTAGTGTATTTATATAT 1020

1021 TAAATGATTTCTGACTCTATCATTTGGCCCTCAAAACACAGATTTGTTTCTTTGGTTTT 1080
1021 TAAATGATTTCTGACTCTATCATTTGGCCCTCAAAACACAGATTTGTTTCTTTGGTTTT 1080

1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCAGCAGTGCCTGCACATAGAAAACAATC 1140
1081 GTTTTCTTCACTATGGGATCTTCTGTGCCAGCAGCAGTGCCTGCACATAGAAAACAATC 1140

1141 AATATTTCTGAATTAATGATTAATAAATCAGAGAACTTTCCCATCTCTGTTGGATCTAT 1200
1141 AATATTTCTGAATTAATGATTAATAAATCAGAGAACTTTCCCATCTCTGTTGGATCTAT 1200

1201 AGAATCATCCAGATTAAGTATGAGGCTCTGCAATTTATATGCGCTTAAATTTAAGATTAT 1260
1201 AGAATCATCCAGATTAAGTATGAGGCTCTGCAATTTATATGCGCTTAAATTTAAGATTAT 1260

1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAATATATAGTAAACACTTTGAAA 1320
1261 GTGAGAAAAGTTTAAAGACACTTAGTAGAGTGATTTTGAATATATAGTAAACACTTTGAAA 1320

1321 TGGTGGTCTTTTAAAGATATTAATAGATTAATGAAAATCTCCATCTCAAAAATAATG 1380
1321 TGGTGGTCTTTTAAAGATATTAATAGATTAATGAAAATCTCCATCTCAAAAATAATG 1380

1381 CATAACTATTTTAAAGGAATCACTCTCAGGCTTTCAATGTTGTTGTTTCTTCTTTT 1440
1381 CATAACTATTTTAAAGGAATCACTCTCAGGCTTTCAATGTTGTTGTTTCTTCTTTT 1440

1441 CATATATTTTACCATCTGCTGAAGGCACTCATATCAAGGGTAAAGAAAGATGGAGGA 1500
1441 CATATATTTTACCATCTGCTGAAGGCACTCATATCAAGGGTAAAGAAAGATGGAGGA 1500

1501 AAACCTCAGTAAGAAATTAATAGTCTGTTTGAAGATGAAAAGATTTCTCATCTCAA 1560
1501 AAACCTCAGTAAGAAATTAATAGTCTGTTTGAAGATGAAAAGATTTCTCATCTCAA 1560

1561 CCTTATGAGCAGGAAGAGGAGGCTGTTTCAGAACCAATTTTACCTAGCAGAACCAATAT 1620
1561 CCTTATGAGCAGGAAGAGGAGGCTGTTTGAAGAACCAATTTTACCTAGCAGAACCAATAT 1620

1621 TTTAGACACTTCCCTGCAATTAACCTGCACAAACAAATATGTTTGCAAAACCTTGTTRGATCAAC 1680
```

Db 1621 TTTAGACATCTCCCTGCAATTAACGTCACAAACAATATGTTGCAAACTGTT-GATCAAC 1679
QY 1681 CTCACAACGACACATTCAGGAGTTAAATATATTTTCAACAACATGGATTTTCTCTTA 1740
Db 1680 CTCACAACGACACATTCAGGAGTTAAATATATTTTCAACAACATGGATTTTCTCTTA 1739
QY 1741 ACCTAGAGATTCTCAAACTCTCTGAAAGGTTCTCAATGGCTTCAGGCTTAAGAAGAT 1800
Db 1740 ACCTAGAGATTCTCAAACTCTCTGAAAGGTTCTCAATGGCTTCAGGCTTAAGAAGAT 1799
QY 1801 TTTCTCCTGTTAATAGCAGCAACAATTTAGCAATTTCACTCTCAAACTTCACATAATGA 1860
Db 1800 TTTCTCCTGTTAATAGCAGCAACAATTTAGCAATTTCACTCTCAAACTTCACATAATGA 1859
QY 1861 TCACATCTTTTCCAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGAACATCAAAA 1920
Db 1860 TCACATCTTTTCCAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGAACATCAAAA 1919
QY 1921 CTAACCATCTGAGAAACTTCCCAAGTGTAAAGACTCTGCCATTTAAACATTAACCGAGA 1978
Db 1920 CTAACCATCTGAGAAACTTCCCAAGTGTAAAGACTCTGCCATTTAAACATTAACATAA 1977

RESULT 5

US-08-692-787-3
; Sequence 3, Application US/08692787
; Patent No. 5882864
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE
; TITLE OF INVENTION: DISEASE
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/692,787
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Corder, Timothy S.
; REGISTRATION NUMBER: 38,414
; REFERENCE/DOCKET NUMBER: UROC:012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 757 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-692-787-3

Query Match 25.0%; Score 626.2; DB 2; Length 757;
Best Local Similarity 94.6%; Pred. No. 9.8e-151;
Matches 716; Conservative 0; Mismatches 28; Indels 13; Gaps 6;
QY 1686 ACAACGACACATTCAGGAGTTAAATATTTTTCATCAAAACATTCGGATTTTCTCTTAACGCT 1745

Db 1 ACAACGACACATTCAGGAGTTAAATATTTATCATCAAACTTGGATTTTCTCTTAACGCT 60
QY 1746 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTTAAGAAGATTTCTC 1805
Db 61 AGAGATTGCTACAAATCTTCTGAAGGGTCTCAATGGCTTCAGGCTTAAGAAGATTTCTC 120
QY 1806 CCGTGTATTAAGCAGCAACAATTTAGCCATTTCACTCTCAAACTTCACATAATGATCACA 1865
Db 121 CCGTGTATTAAGCAGCAACAATTTAGCCATTTCACTCTCAAACTTCACATAATGATCACA 180
QY 1866 TTTCTTTCCAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGAACATCAAACTAAC 1925
Db 181 TTTCTTTCCAAAGGAACCTCTAGAGACCAAAATGCCCGAGTTAAGAACATCAAACTAAC 240
QY 1926 CATCTGAAGAAACTTTCCCAAGTGTAAAGACTCTGCCATTTAAACATTAACGAGGGGACT 1985
Db 241 CATCTGAAGAAACTTTCCCAAGTGTAAAGACTCTGCCATTTAAACATTAACGAGGGGACT 300
QY 1986 CAAACAGTCTTTT - CTTCTCTTTGTCGTGTTTC - TTGCTCTCCAGACCAAGG - ----CACT 2036
Db 301 CAAACAGTCTTTTCTTTCTCTTTGTCGTGTTTC - TTGCTCTCCAGACCAAGG - ----CACT 360
QY 2037 GACGACAGTACTGATACATAATTTAAAGCA - CACTCTCTTCCACTTTGGTAA - -TACCA 2093
Db 361 GACGACTACTTGATACAATAATTTAAAGCACCACCTCTCTTCCCACTTTGTATAATACCA 420
QY 2094 GAACTCTAATTTGGACACCTGGAAGCTTAGGA - CTACGAGCCATACAAATAGTAAACTCT 2152
Db 421 GAACTCTAATTTGGACACCTGGAAGCTTAGGA - CTACGAGCCATACAAATAGTAAACTCT 480
QY 2153 GTCCAGGATTCACTCATCTGTGTATTTTCTATAGATTTTACTAGCGTTTCTGTATATAA 2212
Db 481 GTCCAGGATTCACTCATCTGTGTATTTTCTATAGATTTTACTAGCGTTTCTGTATATAA 540
QY 2213 AAATACCCCGCCAGCGAGCTGAGTGTAACTGTAATCCAGCACTTTGGAGGTGGGT 2272
Db 541 AAATACCCCGCCAGCGAGCTGAGTGTAACTGTAATCCAGCACTTTGGAGGTGGGT 600
QY 2273 GGATCACCCTGAGTCTGGAGTTTCGAGACCAAGCTTCACGAGCATGTGGAAACCCCATCTC 2332
Db 601 GGATCACCCTGAGTCTGGAGTTTCGAGACCAAGCTTCACGAGCATGTGGAAACCCCATCTC 660
QY 2333 TACTAAAAACACAAAAAATTTAGCCGGGTGTGGCAGCATGCTGTATATCCAGTACTC 2392
Db 661 TACTAAAAACACAAAAAATTTAGCCGGGTGTGGCAGCATGCTGTATATCCAGTACTC 720
QY 2393 AGGAGGCTGAGCGGAGAAATTCCTTGAACCCCGGAAGG 2429
Db 721 AGGAGGCTGAGCGGAGAAATTCCTTGAACCCCGGAAGG 757

RESULT 6

US-09-097-199-3
; Sequence 3, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,199
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
FILING DATE: 31-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Nakashima, Richard A.
REGISTRATION NUMBER: P-42,023
REFERENCE/DOCKET NUMBER: UROC:018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 757 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-097-199-3

Query Match 25.0%; Score 626.2; DB 3; Length 757;
Best Local Similarity 94.6%; Pred. No. 9.8e-151;
Matches 716; Conservative 0; Mismatches 28; Indels 13; Gaps 6;

1686 ACACGACACATTCAGGAGTAAATATTTTCATCAAAACATTTGGATTTTTCCTTAACGGT 1745
|||||
1 ACAACGACACATTCAGGAGTAAATATTTTCATCAAAACATTTGGATTTTTCCTTAACGGT 60
|||||

1746 AGAGATTGCTACAAATCTTCTGAAGGCTCAATGCTTTCAGGCTTAAGAGATTTCTC 1805
|||||
61 AGAGATTGCTACAAATCTTCTGAAGGCTCAATGCTTTCAGGCTTAAGAGATTTCTC 120
|||||

1806 CCGTTTATAAGCAGCAAGACAAATAGCCATTTCACTCTCAAACTTCACTAATGATCA 1865
|||||
121 CCGTTTATAAGCAGCAAGACAAATAGCCATTTCACTCTCAAACTTCACTAATGATCA 180
|||||

1866 TTCTTTCAAAGGAAGTCTAGAGACCAATGCCCCGAGTTAAGACATCAAACTAAC 1925
|||||
181 TTCTTTCAAAGGAAGTCTAGAGACCAATGCCCCGAGTTAAGACATCAAACTAAC 240
|||||

1926 CATCTGAAGAACTTCCCAAGTGTAGACTCTGCCATTAATAACATTAACGAGGGGACT 1985
|||||
241 CATCTGAAGAACTTCCCAAGTGTAGACTCTGCCATTAATAACATTAACGAGGGGACT 300
|||||

1986 CAAACAGTCTTTTCTTCCCTTTGTGCTGTTTC-TTGCTCCAGACCAAGG-----CACT 2036
|||||
301 CAAACAGTCTTTTCTTCCCTTTGTGCTGTTTC-TTGCTCCAGACCAAGGACTTGGGG 360
|||||

2037 GAGGACAGTACTGATACATAATTTAAAGCA-CATCCCTTCCACTTTGGTAA--TACCA 2093
|||||
361 GACAGTACTTGATACATAATTTAAAGCAACCACTCCCTTCCACTTTGTAATACCA 420
|||||

2094 GAACTCTAATTGGACCACCTGAAGCTTAGGA-CTACCAAGCCATCAAAATAGTAAACTCT 2152
|||||
421 GAACTCTAATTGGACCACCTGAGCTTAGGACCTACCAAGCCATCAAAATAGTAAACTCT 480
|||||

2153 GTCCAGGATTCATCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTATATA 2212
|||||
481 GTCCAGGATTCATCATCTGTGTATTTTCTATAGATGTTTACTAGGCGTTTGTATATA 540
|||||

2213 AAATACCCCGCCAGGACGGTGGCTCAGCCCTGTAATCCAGCACTTTGGGAGTGGGT 2272
|||||
541 AAATACCCCGCCAGGACGGTGGCTCAGCCCTGTAATCCAGCACTTTGGGAGTGGGT 600
|||||

2273 GGATCACTCTGAGGTGGGAGTTCGAGACCAAGCTGACCAAGCTGAGGAGTGGGAAACCCCATCTC 2332
|||||
601 GGATCACTCTGAGGTGGGAGTTCGAGACCAAGCTGACCAAGCTGAGGAGTGGGAAACCCCATCTC 660
|||||

2333 TACTAAAAACACAAAAAATTAGCCGGCGTGGTGCAATCCCTGTAAATCCAGTACTC 2392
|||||
661 TACTAAAAACACAAAAAATTAGCCGGCGTGGTGCAATCCCTGTAAATCCAGTACTC 720
|||||

2393 AGGAGGCTGAGCGGAGAAATTGCTTGAACCCGGGAGG 2429
|||||
721 AGGAGGCTGAGCGGAGAAATTGCTTGAACCCGGGAGG 757
|||||

RESULT 7
US-09-949-016-119365/c
; Sequence 119365, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: PsetSeq for Windows Version 4.0
; SEQ ID NO 119365
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119365

Query Match 22.8%; Score 572.2; DB 4; Length 601;
Best Local Similarity 98.8%; Pred. No. 6.2e-137;
Matches 596; Conservative 1; Mismatches 4; Indels 2; Gaps 2;

2 ACCTTAAATATATCGAGGTGGCTAATGATGTATATAATTTACAAAAATTTCTTCTAT 61
|||||
601 ACCTTAAATATATCGAGGTGGCTAATGATGTATATAATTTACAAAAATTTCTTCTA- 543
|||||

62 TGCTACAGAGCTACAATTTCAATTTACAGTGCCACCATGAGGGCTTCTTAAAGAACCA 121
|||||
542 TGCTACAGAGCTACAATTTCAATTTACAGTA-GTAAATATGAGGGCTTCTTAAAGAACCA 484
|||||

122 GAAATATGAGGATATGCACAAATTTATTCACATTTTACAGATCAGAAATTTGAGGCACAG 181
|||||
483 GAAATATGAGGATATGCACAAATTTATTCACATTTTACAGATCAGAAATTTGAGGCACAG 424
|||||

182 ATTAAGTAACTTTCCCAAGGCTACCAAGGCATTTAGCTCCAGAAAATGTGCTCTTACCATT 241
|||||
423 ATTAAGTAACTTTCCCAAGGCTACCAAGGCATTTAGCTCCAGAAAATGTGCTCTTACCATT 364
|||||

242 CTGCTACAAGGTATTTTGGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGTT 301
|||||
363 CTGCTACAAGGTATTTTGGAAAAAGAAAAAGTAAAAAGAGTCAAAAGGCAACAGAGTT 304
|||||

302 CATTGATTTTCCATAGAACAGTCAACCATGCAATTTCTCATTATCTTCAGAAAGCATATTTCAC 361
|||||
303 CATTGATTTTCCATAGAACAGTCAACCATGCAATTTCTCATTATCTTCAGAAAGCATATTTCAC 244
|||||

362 GACCAATGAAAGGTTCCCTCAATGAAATGTTTCCCTCATTATCTTCAGAAAGCATATTTCAC 421
|||||
243 GACCAATGAAAGGTTCCCTCAATGAAATGTTTCCCTCATTATCTTCAGAAAGCATATTTCAC 184
|||||

422 ATTGATTTTGCAGTTAACTCAGACCCCTAGGTCTGGAATGCTGTCTTCTACTTATCCAA 481
|||||
183 ATTGATTTTGCAGTTAACTCAGACCCCTAGGTCTGGAATGCTGTCTTCTACTTATCCAA 124
|||||

482 AACTATACATCCACAGATCATATAAATCTCAGCCCTCTCAGAAAGCTTTTCCAGAAAAA 541
|||||
123 AACTATACATCCACAGATCATATAAATCTCAGCCCTCTCAGAAAGCTTTTCCAGAAAAA 64
|||||

Qy	542	TAAAAATGTTGAAAAGGCAATTCTGCTACCAATGACTGTTTAAAGCCAGCCCAAGTA	601
Db	63	TAAAAATGTTTCAAAAAGGCAATTCTGCTACCAATGACTGTTTAAAGCCAGCCCAAGTA	4
Qy	602	GAA	604
Db	3	GAA	1

```

RESULT 8
US-09-949-016-119366/c
; Sequence 119366, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 119366
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119366

```

Query Match	22.8%	Score 572.2	DB 4	Length 601
Best Local Similarity	98.8%	Pred. No. 6.2e-137		
Matches 596	Conservative 1	Mismatches 4	Indels 2	Gaps 2
QY	14	TCGAGGTGGCTAA	TTGATGTATATAAT	TTTACAAAATTTATTTCTTCTATGCTACAGAGCT 73
Db	601	TCGAGGTGGCTAA	TTGATGTATATAAT	TTTACAAAATTTATTTCTTCTA-TGCTACAGAGCT 543
QY	74	ACAATTCAA	TTTACAGTAGGCGCACCATGAGGGCCCTTCTTAAGGAAC	CAGAAAATATGAGGA 133
Db	542	ACAATTCAA	TTTACAGTA-GTAAATATGAGGGCCCTTCTTAAGGAC	CAGAAAATATGAGGA 484
QY	134	TATGCACAA	TATTTATTCACATTTTACAGATCAGAAAAATGAGGCACAGATTAAGTAAC	TTT 193
Db	483	TATGCACAA	TATTTATTCACATTTTACAGATCAGAAAAATGAGGCACAGATTAAGTAAC	TTT 424
QY	194	CCCAAGGCTAC	CAGGCATTTAGCTTCAGAAAATGTGTCTCTTTACCATTC	TGCTACAAAGT 253
Db	423	CCCAAGGCTAC	CAGGCATTTAGCTTCAGAAAATGTGTCTCTTTACCATTC	TGCTACAAAGT 364
QY	254	ATTTTCGAAAAA	AGAAAAAGTAA	AAAAAGCAAGTCAACAGATTCATTGATTATTC 313
Db	363	ATTTTCGAAAAA	AGAAAAAGTAA	AAAAAGCAACAGATTCATTGATTATTC 304
QY	314	CATAGAAC	CACTGCACCATCGAATTCACACCCCTTCAGACACACATTCGACCATGAAGG 373	
Db	303	CAYAGAAC	CACTGCACCATCGAATTCACACCCCTTCAGACACACATTCGACCATGAAGG 244	
QY	374	TTCTCTCAAT	GAAATGTTCCCTCATTATCTTCAGAAAGCCATATTTATTCACATTGACTTTTGCA 433	
Db	243	TTCTCTCAAT	GAAATGTTCCCTCATTATCTTCAGAAAGCCATATTTATTCACATTGACTTTTGCA 184	
QY	434	GTAAACTCAG	ACCCCTAGGCTCGGAATGCTGTCTTCTTACTTATTC	CAAAAACATACATATCC 493
Db	183	GTAAACTCAG	ACCCCTAGGCTCGGAATGCTGTCTTCTTACTTATTC	CAAAAACATACATATCC 124
QY	494	ACAGATCAT	ATAAACTCTCAGCCCTGCTCGAAAGCCCTTTTCAGAAAAAT	AAAAATGGTTG 553

123 ACAGATCATATAAATCTCAGCCCTGTCGCAAGCCTTTCAGAAATAAATAATGGTTG 64
Db
554 AAAAGGCAATTCGTACCAATGACTTTTAAAGCCGACCAAGTAACCTGAACCATTCGAA 613
Qy
63 AAAAGGCAATTCGTACCAATGACTTTTAAAGCCGACCAAGTAACCTGAACCATTCGAA 4
Db
614 CTT 616
Qy
3 CTT 1
Db

```

RESULT 9
US-09-949-016-119367/c
; Sequence 119367, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 119367
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-119367

```

Query Match	15.9%; Score 397.6; DB 4; Length 601;
Best Local Similarity	98.6%; Pred. No. 3.6e-92;
Matches 432; Conservative 1; Mismatches 0; Indels 5; Gaps 3;	
QY	2069 ACTCCCTCCACATTTCGGTAAATACCAAGATCTTAATTGGACCACTTGAAGCTTTAGACTA 2128
Db	
601	ACTCCCTTCACATTTCGTTTATACCAAGATCTTAATTGGACCACTTGAAGCTTTAGACTA 544
QY	2129 CCAGCCATACAAATAGTAAACTCTGTCCAGATTCACTCATCTGTGTATTTTCTATAGAT 2188
Db	
543	CCAGCCATACAAATAGTAAACTCTGTCCAGATTCACTCATCTGTGTATTTTCTATAGAT 484
QY	2189 GTTTTACTAGCGTTTGTTATATAAAATAACCCGGCCAGGCACGGTGGCTCACGCCTGTA 2248
Db	
483	GTTTTACTAGCGTTTGTTATATAAAATAACCCGGCCAGGCACGGTGGCTCACGCCTGTA 424
QY	2249 ATCCGACGACTTTGGGAGTGGGTGGATCACTGAGTGGGAGTTCGAGACAGAGCTGA 2308
Db	
423	ATCCGACGACTTTGGGAGTGGGTGGATCACTGAGTGGGAGTTCGAGACAGAGCTGA 364
QY	2309 CCAGCATGTGGAAACCCCATCTCTACTAAAAACACAAAAATATAGCCGGCGCTGTGGC 2368
Db	
363	CCAGCATGTGGAAACCCCATCTCTACTAAAAACACAAAAATATAGCCGGCGCTGTGGC 304
QY	2369 ACATGCTCTTAATCCGAGCTACTCAGGAGGCTGAGCGGAGAAATTCCTTGAACCCCGAAG 2428
Db	
303	ACRTGCTGTATCCGAGTACTCAGGAGGCTGAGCGGAGAAATTCCTTGAACCCCGAAG 244
QY	2429 GTGAGGTTTGTTCGGTGAAGCTGAGATTGCATTTTGCATCCAGCTCGGGCAACAGGAG 2488
Db	
243	GTGGAG--GTTGCGGTGAGCTGAGATTGCATTTTGCATCCAGCTCGGGCAACAGGAG 187
QY	2489 TAAACTCCCCCCCACC 2506
Db	
186	TAAACTCCCCCCCACC 169

	/	PRIOR APPLICATION NUMBER:	60/231,498	
	/	PRIOR FILING DATE:	2000-09-08	
	/	NUMBER OF SEQ ID NOS:	207012	
	/	SOFTWARE:	FastSeq for Windows Version 4.0	
	/	SEQ ID NO	16509	
	/	LENGTH:	174639	
	/	TYPE:	DNA	
	/	ORGANISM:	Human	
	/	NAME/KEY:	misc feature	
	/	LOCATION:	(1)-(174639)	
	/	OTHER INFORMATION:	n = A,T,C or G	
	/	US-09-949-016-	16509	
		Query Match	7.6%; Score 189.4; DB 4; Length 174639;	
		Best Local Similarity	81.7%; Pred.No. 1.1e-37;	
		Matches	246; Conservative 0; Mismatches 46; Indels 9; Gaps 2;	
Qy		2202	TGTTTATATAAAATACCCGGCAGGCACGGTGGCTCACGCCCTGTATATCCAGCACTTT	2261
Db		112900	TTGTGTACTTTGAAAATGGTGGCGCAGCGGGTGCTCACGCCCTGTATATCCAGCACTTT	112959
Qy		2262	GCGAGGT-----GGGTGGATCACCTCAGGTCGGGAGTTTCGAGACCAGCCTGACCAGCAT	2315
Db		112960	GCGAGGCTGAGCGGGCGGATCACCTCAGGTTGGAGTTTGAGACCAGCCTGATCAACAT	113019
Qy		2316	GGTGGAACCCCCCATCTCTACTTAAAAAACAACAAAAATTAGCGGCGTGGTGCCACATGCC	2375
Db		113020	GGAGAACCCTCGTCTCTACTTAAAAATACAAAAGATTAGCTGGCGTGGTGCCATGCC	113079
Qy		2376	TGTAATCCCAGCTACTCAGAGGCTGAGCGGAGAAATGCTTGAACCCGGAAGGTGGAGG	2435
Db		113080	TGTAATCCCAGCTACGTGGGAGGCTGAGGAGGAGAAATCCCTTGAACCCGGAAGGCGGAG-	113138
Qy		2436	TTGTTGGGTGAGCTGAGATTGCACTATTGCACCTCCAGCCTGGGCCAACAGGAGTAAAACT	2495
Db		113139	--GTTGAGTGAGCCGAGATACGCCATTGCATCTCCAGCTGGGCTACAGAGCGAAACT	113196
Qy		2496	C 2496	
Db		113197	C 113197	
		RESULT 12		
		US-09-949-016-121999		
		/ Sequence 121999, Application US/09949016		
		/ Patent No. 6812339		
		/ GENERAL INFORMATION:		
		/ APPLICANT: VENTER, J. Craig et al.		
		/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED		
		/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF		
		/ FILE REFERENCE: CL001307		
		/ CURRENT APPLICATION NUMBER: US/09/949,016		
		/ CURRENT FILING DATE: 2000-04-14		
		/ PRIOR APPLICATION NUMBER: 60/241,755		
		/ PRIOR FILING DATE: 2000-10-20		
		/ PRIOR APPLICATION NUMBER: 60/237,768		
		/ PRIOR FILING DATE: 2000-10-03		
		/ PRIOR APPLICATION NUMBER: 60/231,498		
		/ PRIOR FILING DATE: 2000-09-08		
		/ NUMBER OF SEQ ID NOS: 207012		
		/ SOFTWARE: FastSeq for Windows Version 4.0		
		/ SEQ ID NO 121999		
		/ LENGTH: 601		
		/ TYPE: DNA		
		/ ORGANISM: Human		
		/ US-09-949-016-121999		
		Query Match	7.5%; Score 189; DB 4; Length 601;	
		Best Local Similarity	84.5%; Pred.No. 1.1e-38;	
		Matches	261; Conservative 1; Mismatches 36; Indels 11; Gaps 4;	
Qy		2202	TGTTTATATAAAATACCCGGCAGGCACGGTGGCTCACGCCCTGTATATCCAGCACTTT	2261

Db 105 TTTTATAGATGCTAAGCAGGCGAGCGGCTGAGCGCTATAATCCAGCACTTT 164
QY 2262 GGG-----AGGTGGGTGATACCTCAGCTCGGAGTTCGAGACCAGGCTGACCAAGCAT 2315
Db 165 GGGAGGTCAAGGTGGCGAGATCACTGAGTCAAGGTTCGAGACCAGGCTGACCAACAT 224
QY 2316 GGTGGAACCCCATCTCTACTATAAAACACAAAAATAGCCGGGCTGTGGGCATGCC 2375
Db 225 GG-AGAAACCCCAACTCTACTAAAAATACAAAAATTAGCCAGGCATGTGGGCATGCC 283
QY 2376 TGTAATCCAGCTACTCAGGAGCTGAGGCGAGAGATTGCTTGAACCCGGAGGTGGAG 2434
Db 284 TGTAATCCAGCTACTCAGGAGCTGAGGCGAGAGATTGCTTGAACCCGGAGGTGGAG 343
QY 2435 GTTGTGCGGTGAGCTGAGATTGCACTATTGCACTCCAGCTGGGCAACAGGAGTAAAC 2494
Db 344 ---GTTGCGGTGAGCTGAGATGAGCGCACTTGCATCCAGTCTGGCGCAACAGGACAAAC 400
QY 2495 TCCCCCCCCA 2503
Db 401 TCTGCACCA 409

RESULT 13
US-09-949-016-16308
; Sequence 16308, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16308
; LENGTH: 48691
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(48691)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16308

Query Match 7.5%; Score 188; DB 4; Length 48691;
Best Local Similarity 83.4%; Pred. No. 1.5e-37;
Matches 252; Conservative 0; Mismatches 40; Indels 10; Gaps 3;
QY 2209 ATAAAAATACCCCGGCAGGCGGTGGCTCAGCGCTGTAATCCAGCACTTTGGGAGGT 2268
Db 42100 AAAAAATAAACAGGCGGAGTATGGTGGCTCAGCGCTGTAATCCAGCACTTTGGGAGGC 42159
QY 2269 -----GGGTGGATCACTGAGTCCGGAGTTCGAGACCAGGCTGACAGCATGGTGGAA 2322
Db 42160 CGAGGGGGTGGATCACTGAGTCAAGGATTTGAGACAGGCTGGCCCAACATGGTGAAC 42219
QY 2323 CCCCCATCTCTACTATAAACACAAAAATAGCCGGGCTGGTGGGCATGCTGTAATC 2382
Db 42220 CCCCCATCTCTACTATAAACACAAAAATAGCCAGATGGTGGTGCAGCGCTGTAATC 42279
QY 2383 CCAGCTACTCAGGAGGTGAGGC- GGAGAAATTGCTTGAACCCCGAAGGTGGAGTTGTTG 2441
Db 42280 CCAGCTACTTGGAGGCTGAGGAGGAGATCGCTTGAACCCCGGAGGCGGAG---GTTG 42336

QY 2442 CGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAACTCCCCC 2501
Db 42337 CAGTGGAGCTGAGATCGCACCCTCAGCTCCAGCTGGATGACAAGAGTGAATCCGTCT 42396
QY 2502 CA 2503
Db 42397 CA 42398

RESULT 14
US-09-949-016-15165/c
; Sequence 15165, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15165
; LENGTH: 70947
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15165

Query Match 7.5%; Score 187.8; DB 4; Length 70947;
Best Local Similarity 84.5%; Pred. No. 1.9e-37;
Matches 261; Conservative 0; Mismatches 37; Indels 11; Gaps 4;
QY 2202 TTGTATATATAAAATACCCCGGCAGGCGGTGGCTCAGCGCTGTAATCCAGCACTTT 2261
Db 47525 TTGTTAGGAATGCTTAAGCAGGCGGAGGCGGTGGCTGAGCTATAATCCAGCACTTT 47466
QY 2262 GGG-----AGGTGGGTGATCACTGAGTCCGAGTTCGAGACCAGGCTGACCAAGCAT 2315
Db 47465 GGGAGGTCAAGGTGGGCGAGATCACCTGAGTCAAGGTTCGAGACCAGGCTGACCAACAT 47406
QY 2316 GGTGGAACCCCATCTCTACTATAAAACACAAAAATAGCCGGGCTGGTGGCAGATGCC 2375
Db 47405 GG-AGAAACCCCACTCTACTATAAAATACAAAAATAGCCAGGATGGTGGCAGATGCC 47347
QY 2376 TGTAAATCCAGCTACTCAGGAGGCTGAGGC- GGAGAAATTGCTTGAACCCGGAAGGTGGAG 2434
Db 47346 TGTAAATCCAGCTACTCAGGAGGCTGAGGCAGGAGAAATGCTTGAACCCGGGAGGAG 47287
QY 2435 GTTGTGGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAC 2494
Db 47286 ---GTTGGGTGAGCTGAGATGACGCCATTCGACTCCAGTCTCTGGGCAACAGGACAAAC 47230
QY 2495 TCCCCCCCCA 2503
Db 47229 TCTGCACCA 47221

RESULT 15
US-09-949-016-163069/c
; Sequence 163069, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 163069
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-163069

Query Match      7.5%; Score 187.6; DB 4; Length 601;
Best Local Similarity 83.1%; Pred.No.2.4e-38;
Matches 251; Conservative 1; Mismatches 40; Indels 10; Gaps 3;

QY 2209 ATAAAAATACCCGCGCAGGCGAGTGGCTCAGCGCTGTAATCCAGCACTTTGGGAGGT 2268
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
522 AAAAAATAAACAGGCCGAGTATGGTGGCTCAGCGCTGTAATCCAGCACTTTGGGAGGC 463
QY 2269 -----GGGTGGATCACCTGAGGTGGGAGTTGAGACCAGCGCTGGCCAAACATGGTGAAA 2322
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 CGAGCGCGGTGGATCACCTGAGGTGAGAGTTTGAGACCAGCGCTGGCCAAACATGGTGAAA 403
QY 2323 CCCCCATCTCTACTAAAAACACAAAAAATTAGCCGGCGTGGTGGCACATGCTGTAAATC 2382
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 CCCCCATCTCTACTAAAAATACAAAAAATTAGCCAGATGTGGTGGTCACGCCCTGTAAATC 343
QY 2383 CCAGCTACTCTCAGGAGGCTGAGGC-GGAGAAATTGCTTTGAACCCGGAAGGTGGAGTTGTTG 2441
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 CCAGCTACTTTGGAGGCTCAGGCAGGAGAAATCGCTTGAACCCYGGGAGGCGGAG---GTTG 286
QY 2442 CGGTGAGCTGAGATTGCACTATTGCACTCCAGCCTGGGCAACAGGAGTAAAACTCCCCCC 2501
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
285 CAGTGAGCTGAGATCGCACCTGCACTCCAGCCTGGATGACAAGAGTGAACCTCCGTCT 226
QY 2502 CA 2503
Db      |||
225 CA 224
```

Search completed: September 9, 2005, 06:29:07
Job time : 378.657 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 9, 2005, 03:30:51 ; Search time 1331.01 Seconds
(without alignments)
12364.714 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gacctaataatcgaggt.....agtaaaactccccccacc 2506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2494	99.5	2505	10	US-09-974-546-85 Sequence 85, Appl
2	1950	77.8	2087	10	US-09-974-546-83 Sequence 83, Appl
3	626.2	25.0	757	10	US-09-974-546-3 Sequence 3, Appli
C 4	190.6	7.6	299598	19	US-10-322-696-16 Sequence 16, Appl
C 5	189.4	7.6	129447	20	US-10-719-993-6897 Sequence 6897, Ap
6	186.4	7.4	14417	9	US-09-860-670-251 Sequence 251, App
7	186.4	7.4	14417	17	US-10-227-646-251 Sequence 251, App

8	186.4	7.4	14426	9	US-09-860-670-249 Sequence 249, App
9	186.4	7.4	14426	17	US-10-227-646-249 Sequence 249, App
C 10	186.2	7.4	532	13	US-10-027-632-187685 Sequence 187685,
C 11	186.2	7.4	532	13	US-10-027-632-187686 Sequence 187686,
C 12	186.2	7.4	532	17	US-10-027-632-187685 Sequence 187685,
C 13	186.2	7.4	532	17	US-10-027-632-187686 Sequence 187686,
C 14	186.2	7.4	43623	21	US-10-741-600-17818 Sequence 17818, A
15	184.6	7.4	450	13	US-10-027-632-266585 Sequence 266585,
16	184.6	7.4	450	17	US-10-027-632-266585 Sequence 266585,
17	184.4	7.4	600	21	US-10-956-157-7352 Sequence 7352, Ap
18	184.4	7.4	1449	21	US-10-887-553A-909 Sequence 909, App
19	184.4	7.4	1449	21	US-10-956-157-2117 Sequence 2117, Ap
20	184.4	7.4	1599	14	US-10-153-668-439 Sequence 439, App
C 21	184.4	7.4	3079	13	US-10-027-632-112690 Sequence 112690,
C 22	184.4	7.4	3079	13	US-10-027-632-112691 Sequence 112691,
C 23	184.4	7.4	3079	17	US-10-027-632-112690 Sequence 112690,
C 24	184.4	7.4	3079	17	US-10-027-632-112691 Sequence 112691,
25	183.8	7.3	44348	19	US-10-301-832-11 Sequence 11, Appl
26	183.6	7.3	106746	20	US-10-856-888-12 Sequence 12, Appl
27	183.6	7.3	106746	20	US-10-856-888-12 Sequence 12, Appl
C 28	183.6	7.3	108317	15	US-10-292-798-1789 Sequence 1789, Ap
C 29	183.6	7.3	108317	15	US-10-017-161-2143 Sequence 2143, Ap
C 30	183.4	7.3	585	13	US-10-027-632-106412 Sequence 106412,
C 31	183.4	7.3	585	17	US-10-027-632-106412 Sequence 106412,
C 32	183.4	7.3	723	13	US-10-027-632-22469 Sequence 22469, A
33	183.4	7.3	723	17	US-10-027-632-22469 Sequence 22469, A
34	183.2	7.3	110665	20	US-10-723-860-2002 Sequence 2002, Ap
35	183.2	7.3	110665	22	US-10-756-149-1888 Sequence 1888, Ap
36	183.2	7.3	227968	20	US-10-723-860-1357 Sequence 1357, Ap
C 37	183	7.3	73723	19	US-10-322-281-118 Sequence 118, App
38	182.8	7.3	463	10	US-09-918-995-23846 Sequence 23846, A
39	182.6	7.3	89625	13	US-10-087-192-2020 Sequence 2020, Ap
40	182.6	7.3	90614	19	US-10-331-053-64 Sequence 64, Appl
C 41	182.2	7.3	71048	13	US-10-087-192-352 Sequence 352, App
C 42	182	7.3	16156	15	US-10-017-161-981 Sequence 981, App
C 43	181.8	7.3	91760	13	US-10-087-192-844 Sequence 844, App
44	181.6	7.2	14426	9	US-09-860-670-252 Sequence 252, App
45	181.6	7.2	14426	17	US-10-227-646-252 Sequence 252, App

ALIGNMENTS

RESULT 1
US-09-974-546-85
; Sequence 85, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert

TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS, PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/974,546
FILING DATE: 10-Oct-2001
CLASSIFICATION: Unknown
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/097,199

/ FILING DATE: 1998-06-12
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 85:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2505 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 99..503
/ SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-974-546-85

Query Match 99.5%; Score 2494; DB 10; Length 2505;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2505; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	1	GACCTTAAATATATCGAGGTGGCTAATTGATGTATATAATTTACAAAATTTCTTCTTA	60
DB	1	GACCTTAAATATATCGAGGTGGCTAATTGATGTATATAATTTACAAAATTTCTTCTTA	60
QY	61	TTGCTACAGAGCTACAAATTCATTTACAGTAGGCCACCATGAGGGCTTTCTTAAGGAACC	120
DB	61	TTGCTACAGAGCTACAAATTCATTTACAGTAGGCCACCATGAGGGCTTTCTTAAGGAACC	120
QY	121	AGAAATATGAGGATATGCAATATATTATTTACATTTTACAGATCAGAAAATTTGAGGCACA	180
DB	121	AGAAATATGAGGATATGCAATATATTATTTACATTTTACAGATCAGAAAATTTGAGGCACA	180
QY	181	GATTAAGTAATCTCCCAAGGCTACCAAGGCTTCTAGCTCCAGAAATCTGTGCTTTACCAT	240
DB	181	GATTAAGTAATCTCCCAAGGCTACCAAGGCTTCTAGCTCCAGAAATCTGTGCTTTACCAT	240
QY	241	CTGCTACAAAGGTATTTCCAAAAAGAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT	300
DB	241	CTGCTACAAAGGTATTTCCAAAAAGAAAAGTAAAAAGAGTCAAAAGGCAACAGAGT	300
QY	301	TCATTGATTTATCCATAGAACAGTACACCATGCAATTTCTCACCCCTTTGCGACACACT	360
DB	301	TCATTGATTTATCCATAGAACAGTACACCATGCAATTTCTCACCCCTTTGCGACACACT	360
QY	361	TGACCATGAAAGGTTCTCAATGAAATGTTCTCTCAATTTCTTTCAGAACCATATTATCCA	420
DB	361	TGACCATGAAAGGTTCTCAATGAAATGTTCTCTCAATTTCTTTCAGAACCATATTATCCA	420
QY	421	CATTGACCTTTGAGTTAACTCAGACCCCTAGGCTGGAATGCTGTCTCTACTATTATCCA	480
DB	421	CATTGACCTTTGAGTTAACTCAGACCCCTAGGCTGGAATGCTGTCTCTACTATTATCCA	480
QY	481	AAACTATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA	540
DB	481	AAACTATACATCCACAGATCATATAAATCTCAGCCCTGCTGCAAGCCCTTTCCAGAAA	540
QY	541	ATAAAAATGGTTGAAAGCAATTTCTGCTACCAATGCAATGCTTTTAAGCCCGAGCAAGTAA	600
DB	541	ATAAAAATGGTTGAAAGCAATTTCTGCTACCAATGCAATGCTTTTAAGCCCGAGCAAGTAA	600
QY	601	TGACCATTTCCAACTTCAATTTTATGAAAGAAATTTGATGATGAGGTTATTTTC	660
DB	601	TGACCATTTCCAACTTCAATTTTATGAAAGAAATTTGATGATGAGGTTATTTTC	660
QY	661	AATTTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAATCATAGATTTATCTCT	720
DB	661	AATTTCTAAAATACAAACCCATGTTGATCTTTCTCAATCTTGAATCATAGATTTATCTCT	720
QY	721	ATTATCTCAATTTAGTTTGTATTATTCCTAGTGGGCCATTTAAAACTACCAATGTTGTT	780

DB	721	ATTATCTCAATTTAGTTTGTATTATTCCTAGTGGCCATTTAAAAAATACCAATGTTGTT	780
QY	781	TCGTCTCTCCATTTAGTCAATAACTAAACTAACAGCAATTTAGTAGCCATGTGCCAGAT	840
DB	781	TCGTCTCTCCATTTAGTCAATAACTAAACTAACAGCAATTTAGTAGCCATGTGCCAGAT	840
QY	841	GCTCCGCTAGGCACACAGAGGGATAAAAACAATPACTTTATAGTATACCACTTAATTTTCGCTT	900
DB	841	GCTCCGCTAGGCACACAGAGGGATAAAAACAATPACTTTATAGTATACCACTTAATTTTCGCTT	900
QY	901	AGTAACTAGTGAATGTTCAAGTCAATGCTGAGTCAAGAGTTTGAGGAGACATTACAATGT	960
DB	901	AGTAACTAGTGAATGTTCAAGTCAATGCTGAGTCAAGAGTTTGAGGAGACATTACAATGT	960
QY	961	GTAATGGAAACCAAGAAAAGTGAAACTTTGGTAAGTGGGACTAGTGTATTATATATT	1020
DB	961	GTAATGGAAACCAAGAAAAGTGAAACTTTGGTAAGTGGGACTAGTGTATTATATATT	1020
QY	1021	TAATTTGATTTCTGACTCTATCATTTGGCCTCCAAAACACAGATTTGTTTCTTTTCTTTT	1080
DB	1021	TAATTTGATTTCTGACTCTATCATTTGGCCTCCAAAACACAGATTTGTTTCTTTTCTTTT	1080
QY	1081	GTCTTCTTCACTATGCGGATCTTCTGTGCCAGCACAGTGCCTGCACATAGAAAACAATC	1140
DB	1081	GTCTTCTTCACTATGCGGATCTTCTGTGCCAGCACAGTGCCTGCACATAGAAAACAATC	1140
QY	1141	RATATTTCTGTAATTAATTAATAAATCAGAGAACTTTCCCACTTCTGTTTGGATCTAT	1200
DB	1141	RATATTTCTGTAATTAATTAATAAATCAGAGAACTTTCCCACTTCTGTTTGGATCTAT	1200
QY	1201	AGAACTACAGAGTAAGTGATGAGGGCTCTGCAATTTATATGCGCTTAAATTAAGATTTAT	1260
DB	1201	AGAACTACAGAGTAAGTGATGAGGGCTCTGCAATTTATATGCGCTTAAATTAAGATTTAT	1260
QY	1261	GTGAGAAAGTTTAAAGACACTTTAGTAGAGTGATTTGAAATATATAGTAAACAATTGGAAA	1320
DB	1261	GTGAGAAAGTTTAAAGACACTTTAGTAGAGTGATTTGAAATATATAGTAAACAATTGGAAA	1320
QY	1321	TGGTGGTCTTTAAAGAGATTAATAGATAATTAATAATGAAATCTCCATCTCAAAAATATG	1380
DB	1321	TGGTGGTCTTTAAAGAGATTAATAGATAATTAATAATGAAATCTCCATCTCAAAAATATG	1380
QY	1381	CATAAATCTATTAAAGGAAATCACTCTCAGAGCTTTCAATGTTTGTTCATTACTTTT	1440
DB	1381	CATAAATCTATTAAAGGAAATCACTCTCAGAGCTTTCAATGTTTGTTCATTACTTTT	1440
QY	1441	CATATATTTTACCATCTGCTGAAGGAGTCAATATCAAGGGTAAAGAAAGATGGAGGA	1500
DB	1441	CATATATTTTACCATCTGCTGAAGGAGTCAATATCAAGGGTAAAGAAAGATGGAGGA	1500
QY	1501	AAACTCAGTAAAGATTAATTTAGTCTGTTGCAAGTAAAGAAAGATTTCTCATCTCAA	1560
DB	1501	AAACTCAGTAAAGATTAATTTAGTCTGTTGCAAGTAAAGAAAGATTTCTCATCTCAA	1560
QY	1561	CCTTATGAGCAGGAGGAGGAGGCTGTTTGAGAACCTTTTACTTTAGCAGAACCAATAT	1620
DB	1561	CCTTATGAGCAGGAGGAGGAGGCTGTTTGAGAACCTTTTACTTTAGCAGAACCAATAT	1620
QY	1621	TTTAGACACTTCCCTGCAATTAATCTGCAAAACAATATGTTTGCAAACTTTGTTGATCAAC	1680
DB	1621	TTTAGACACTTCCCTGCAATTAATCTGCAAAACAATATGTTTGCAAACTTTGTTGATCAAC	1680
QY	1681	CTCCAAACACGACACTTTGAGGTTAAATATTTTTCATCAAACTTCGATTTTCTTTT	1740
DB	1681	CTCCAAACACGACACTTTGAGGTTAAATATTTTTCATCAAACTTCGATTTTCTTTT	1740
QY	1741	AGCTTAGAGATTGCTTACAAATCTTCTGAGGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT	1800
DB	1741	AGCTTAGAGATTGCTTACAAATCTTCTGAGGGCTCTCAATGGCTTCAGGCTTAAGAGAGAT	1800
QY	1801	TTCTCCCTGTTTAAAGCAGCAAGCAAAATTTAGCCATTTCTCTCAAACTTCCTAAATGA	1860
DB	1801	TTCTCCCTGTTTAAAGCAGCAAGCAAAATTTAGCCATTTCTCTCAAACTTCCTAAATGA	1860

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2005, 22:58:29 ; Search time 5412.94 Seconds
(without alignments)
17622.418 Million cell updates/sec

Title: US-09-974-546C-85
Perfect score: 2506
Sequence: 1 gacctaataatagaggt.....agtaaaactccccccacc 2506

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	198.8	7.9	700	9 AG163691	Pan trogl
C 2	194.2	7.7	490	8 AQ416174	RPCI-11-1
C 3	187.4	7.5	586	6 CB217256	NISC nb01
C 4	186.4	7.4	487	7 CN275741	170006000
C 5	185.8	7.4	754	7 CN264392	170004243
C 6	185.8	7.4	818	6 CD104583	AGENCOURT
C 7	185.2	7.4	390	7 CN272500	170006000
C 8	184.8	7.4	402	5 BQ631269	in4906.x
C 9	184.8	7.4	472	5 BU790005	in4906.x
C 10	184.4	7.4	344	1 AI590404	ts10b03.x
C 11	184.4	7.4	541	5 BM993720	UI-H-DHO-
C 12	184.4	7.4	613	5 BM990533	UI-H-DHO-
C 13	184.4	7.4	733	6 CA308850	UI-H-FTL-
C 14	184.4	7.4	902	2 BF034649	60145087
C 15	184.4	7.4	906	5 BQ227117	AGENCOURT
C 16	184.4	7.4	949	5 BQ950110	AGENCOURT
C 17	183.8	7.3	918	2 BF526838	602070394
C 18	183.2	7.3	686	3 AG091944	Pan trogl
C 19	183.2	7.3	2900	9 AL832694	Homo sapi
C 20	182.6	7.3	521	5 BQ021987	UI-1-BB1p
C 21	182.2	7.3	969	3 CR623410	full-length
C 22	182	7.3	342	1 AJ573255	AJ573255
C 23	181.6	7.2	357	1 AI031759	ow39dl1.x
C 24	181.6	7.2	366	5 BU736145	UI-E-DX0-

C 25	181.4	7.2	826	5 BQ220083	AGENCOURT
C 26	181.2	7.2	654	9 AQ051672	Pan trogl
C 27	181	7.2	627	1 AV763460	AV763460
C 28	181	7.2	628	6 CA306131	UI-H-FTL-
C 29	181	7.2	668	7 CK003906	AGENCOURT
C 30	181	7.2	716	5 BM969125	UI-CF-ENO
C 31	180.8	7.2	681	5 BQ016777	UI-H-DHO-
C 32	180.8	7.2	862	4 BG762519	602733970
C 33	180.8	7.2	876	6 CD556766	AGENCOURT
C 34	180.8	7.2	909	4 BG034698	602300050
C 35	180.8	7.2	3128	3 CR627380	Homo sapi
C 36	180.8	7.2	3243	3 HSM805448	AL834376
C 37	180.6	7.2	346	5 BM991496	UI-H-DHO-
C 38	180.6	7.2	582	5 BP250579	BP250579
C 39	180	7.2	392	1 AI311796	Q091105.x
C 40	180	7.2	887	5 BQ423685	AGENCOURT
C 41	180	7.2	941	5 BX451504	EX451504
C 42	179.8	7.2	487	6 CD106585	AGENCOURT
C 43	179.8	7.2	678	6 CB270171	1009078 H
C 44	179.8	7.2	679	2 BE392753	601307819
C 45	179.8	7.2	694	6 CA420398	UI-H-FG0-

ALIGNMENTS

RESULT 1
AG163691/c
LOCUS Pan troglodytes DNA, clone: RP43-030103.TJ, genomic survey
DEFINITION sequence.
ACCESSION AG163691
VERSION AG163691.1 GI:16693369
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 700)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpes@gs.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the Red process and may have higher chance of clone tracking errors.

PRIMERS
Sequencing: TJ
LIBRARY Vector : pBACe3.6
R.Site 1 : ECORI
R.Site 2 : ECORI.
Location/Qualifiers
1. .700
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-030103.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"

Query Match 7.9%; Score 198.8; DB 9; Length 700;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: September 9, 2005, 13:07:44 ; Search time 2414 Seconds
(without alignments)
2709.799 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQKVEDMNIHIL.....TLGLECLLYLSKTHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-Q=/cgn2.1/USPTO.spool/US0974546/runat_07092005.174502.20945/app.query.fasta_1.654
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DICALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0974546 @CCN 1.1.2647 @runat_07092005.174502.20945 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*

AR146835
Sequence 83 from patent US 6218529.
AR146835
GI:15110024
Unknown.
Unknown.
Unclassified.
1 (bases 1 to 2087)
An.G., O'Hara.S.Mark., Ralph.D. and Veltri.R.
Biomarkers and targets for diagnosis, prognosis and management of
prostate, breast and bladder cancer
Patent: US 6218529-A 83 17-APR-2001;
Location/Qualifiers
1..2087
/organism="unknown"
/mol_type="unassigned DNA"

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	2087	6 AR146835	Sequence
2	694	100.0	2103	9 AF189270	Homo sapi
3	694	100.0	2496	9 BC069109	Homo sapi
4	694	100.0	2505	6 AR146836	Sequence

5	694	100.0	2512	9 AF189269	Homo sapi
6	694	100.0	129227	9 HSI71M11	Human DNA
7	694	100.0	186698	2 AC036236	Homo sapi
8	99.5	14.3	259320	2 AC096345	Rattus no
9	95	13.7	9833	9 AF291597	Homo sapi
10	94	13.5	9432	9 AF291608	Homo sapi
11	94	13.5	9817	9 AF291631	Homo sapi
12	94	13.5	9817	9 AF291632	Homo sapi
13	94	13.5	9830	9 AF291644	Homo sapi
14	94	13.5	9833	9 AF291587	Homo sapi
15	94	13.5	9833	9 AF291588	Homo sapi
16	94	13.5	9833	9 AF291589	Homo sapi
17	94	13.5	9833	9 AF291590	Homo sapi
18	94	13.5	9833	9 AF291591	Homo sapi
19	94	13.5	9833	9 AF291592	Homo sapi
20	94	13.5	9833	9 AF291593	Homo sapi
21	94	13.5	9833	9 AF291594	Homo sapi
22	94	13.5	9833	9 AF291595	Homo sapi
23	94	13.5	9833	9 AF291598	Homo sapi
24	94	13.5	9833	9 AF291599	Homo sapi
25	94	13.5	9833	9 AF291600	Homo sapi
26	94	13.5	9833	9 AF291601	Homo sapi
27	94	13.5	9833	9 AF291602	Homo sapi
28	94	13.5	9833	9 AF291603	Homo sapi
29	94	13.5	9833	9 AF291604	Homo sapi
30	94	13.5	9833	9 AF291605	Homo sapi
31	94	13.5	9833	9 AF291606	Homo sapi
32	94	13.5	9833	9 AF291607	Homo sapi
33	94	13.5	9833	9 AF291609	Homo sapi
34	94	13.5	9833	9 AF291610	Homo sapi
35	94	13.5	9833	9 AF291612	Homo sapi
36	94	13.5	9833	9 AF291614	Homo sapi
37	94	13.5	9833	9 AF291615	Homo sapi
38	94	13.5	9833	9 AF291618	Homo sapi
39	94	13.5	9833	9 AF291620	Homo sapi
40	94	13.5	9833	9 AF291621	Homo sapi
41	94	13.5	9833	9 AF291623	Homo sapi
42	94	13.5	9833	9 AF291624	Homo sapi
43	94	13.5	9833	9 AF291625	Homo sapi
44	94	13.5	9833	9 AF291626	Homo sapi
45	94	13.5	9833	9 AF291627	Homo sapi

ALIGNMENTS

Alignment Scores:	2.18e-67	Length:	2087
Pred. No.:	694.00	Matches:	135
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	6		

```
US-09-974-546C-86 (1-135) x AR146835 (1-2087)
QY 1 MetArgAlaPheLeuArgAsnGlnIysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCCTCTTAAGGAACAGAAATATGAGGATATGCACAAATATTATTACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAAAATTGAGGCACAGATTAAAGTAACTCCCAAGGCTACACAGCATTTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60
Db 219 CCAGAAACTGTGCTCTTACCATTTGCTACAAGGTATTTCGAAAAAAGAAAAAGTAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 279 AGAAGTCAAAAGGCAACAGATTCATTGATTATTCATAGACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerMetLysCysSerLeu 100
Db 339 CTCACACCCCTTGACAGACACATTCACCATGAAAGGTTCTCTCAATGAATGTTCCCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 399 TCTTCAGAAAGCCATATTATTACATTGACTTTGCAGTTAACTCAGACCCCTAGGCTCTGGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 459 TGTGTGCTCTCTACTTATCCAAACTATACATCCACAGATCATA 503

RESULT 2
AF189270 2103 bp mRNA linear PRI 22-MAR-2001
LOCUS
DEFINITION Homo sapiens UC28 protein (UC28) mRNA, complete cds.
ACCESSION AF189270
VERSION AF189270.1 GI:10441603
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
An.G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
Wright,G.L. Jr. and Veltri,R.W.
Cloning and characterization of UROC28, a novel gene overexpressed
in prostate, breast, and bladder cancers
Cancer Res. 60 (24), 7014-7020 (2000)
21028101
11156405
PUBMED
An.G. and Veltri,R.W.
AUTHORS
Direct Submission
TITLE Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
Parkway, Oklahoma City, OK 73104, USA
FEATURES
Location/Qualifiers
1..2103
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosomes="6"
/map="6q23-q24"
/tissue.type="prostate cancer"
1..2103
/gene="UC28"
99..506
/gene="UC28"
/codon_start=1
/product="UC28 protein"
/protein_id="AG17118.1"
/db_xref="GI:10441604"
/translation="MRAFLRNQKYEEDMNIHILQIRKLRLHSLNFPRLPGILAPETY
LLPFCYKVRKKEKVRKSQKATEFIDYSIEQSHHAILTPLOTHLTMKSSMKCSSLS
```

EAILFTLTQLTQTLGLECCLLYLSKTHPQII"

1961..2103

/gene="UC28"

/note="alternative splicing"

2047..2052

/gene="UC28"

ORIGIN

Alignment Scores:

Pred. No.:	2,2e-67	Length:	2103
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-974-546C-86 (1-135) x AF189270 (1-2103)

QY 1 MetArgAlaPheLeuArgAsnGlnIysTyrGluAspMetHisAsnIleIleHisIleLeu 20

Db 99 ATGAGGGCCCTCTTAAGGAACAGAAATATGAGGATATGCACAAATATTATTACATTTTA 158

QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

Db 159 CAGATCAGAAAAATTGAGGCACAGATTAAAGTAACTCCCAAGGCTACACAGGCATTTCTAGCT 218

QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60

Db 219 CCAGAAACTGTGCTCTTACCATTTGCTACAAGGTATTTCGAAAAAAGAAAAAGTAAA 278

QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80

Db 279 AGAAGTCAAAAGGCAACAGATTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338

QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

Db 339 CTCACACCCCTTGACAGACACATTCACCATGAAAGGTTCTCTCAATGAATGTTCCCTCATTA 398

QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Db 399 TCTTCAGAAAGCCATATTATTACATTGACTTTGCAGTTAACTCAGACCCCTAGGCTCTGGAA 458

QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135

Db 459 TGTGTGCTCTCTACTTATCCAAACTATACATCCACAGATCATA 503

RESULT 3

BC069109

LOCUS

DEFINITION Homo sapiens prostate and breast cancer overexpressed 1, mRNA (cdna

clone IMAGE:7216926).

ACCESSION BC069109

VERSION BC069109.1

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2496)

Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

Hopkins,R.F., Jordan,H., Moore,T., Max,I., Wang,J., Hsieh,F.,

Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,

Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,

Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,

Carninci,P., Prange,C., Raha,S., Loquellano,N.A., Peters,G.J.,

Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,

McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,

Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,

Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,

Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S.,

Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smal, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 2496)
 Strausberg, R.
 Direct Submission
 Submitted (16-APR-2004) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-rc@mail.nih.gov
 Tissue Procurement: Anup Madan, University of Iowa
 CDNA Library Preparation: Anup Madan, University of Iowa
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Neurogenomics Research Lab,
 200 B ENRB, University of Iowa, Iowa City, IA-52242
 anup-madan@uiowa.edu
 Jessica Fahey, Tim Nelson, Jae Goon Yoon and Anup Madan

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: Plate: Row: Column: 0.

FEATURES

Location/Qualifiers
 1..2496
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:7216926"
 /tissue type="Synthetic constructs"
 /clone_lib="NIH_MGC_242"
 /lab_host="XL10"
 /notes="Vector: pUC19"

ORIGIN

Alignment Scores:
 Pred. No.: 2,666-67 Length: 2496
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-09-974-546c-86 (1-135) x BC069109 (1-2496)
 QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
 Db 96 ATGAGGGCCCTCTTAAGAACCCAGAAATATGAGATATGCACATATTATTCACATTTTA 155
 QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 Db 156 CAGATCAGAAAATTGAGGCACAGATTAAAGTAACTTCCCAAGGCTACGAGCATCTTAGCT 215
 QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGlyLysValLys 60
 Db 216 CCAGAAACTGTGCTCTTACCATTTCTGCTACAAAGTATTTTCGAAAAAAGAAAAA 275
 QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGlnSerHisAlaIle 80
 Db 276 AGAAGTCAAAAGGCAACAGAGTTTCATTTATTCATAGAACAGTCACACCATGCATTT 335
 QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
 Db 336 CTCACACCCCTTGACAGACACACTTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 395
 QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

Db 396 TCTTCAGAGCCATATTATTACATTTGACTTTGACAGTTAACTCAGACCTAGGTCTGGAA 455
 QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
 Db 456 TGCTGTCTTCTCTACTATTATCCAAAATATACATATCCACAGATCATTA 500
 RESULT 4
 ARI46836
 LOCUS ARI46836 2505 bp DNA linear PAT 08-AUG-2001
 DEFINITION Sequence 85 from patent US 6218529.
 ACCESSION ARI46836
 VERSION ARI46836.1 GI:15110025
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 2505)
 AUTHORS An, G., O'Hara, S., Mark, R., Ralph, D. and Veltri, R.
 TITLE Biomarkers and targets for diagnosis, prognosis and management of
 prostate, breast and bladder cancer
 JOURNAL Patent: US 6218529-A 85 17-APR-2001;
 FEATURES Location/Qualifiers
 1..2505
 /organism="unknown"
 /mol_type="unassigned DNA"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,686-67 Length: 2505
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-974-546c-86 (1-135) x ARI46836 (1-2505)
 QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
 Db 99 ATGAGGGCCCTCTTAAGGAACCCAGAAATATGAGGATATGCACATATTATTCACATTTTA 158
 QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 Db 159 CAGATCAGAAAATTGAGGCACAGATTAAAGTAACTTCCCAAGGCTACGAGCATTTAGCT 218
 QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGlyLysValLys 60
 Db 219 CCAGAAACTGTGCTCTTACCATTTCTGCTACAAAGTATTTTCGAAAAAAGAAAAA 278
 QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGlnSerHisAlaIle 80
 Db 279 AGAAGTCAAAAGGCAACAGAGTTTCATTTATTCATAGAACAGTCACACCATGCATTT 338
 QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
 Db 339 CTCACACCCCTTGACAGACACACTTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 398
 QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 Db 399 TCTTCAGAGCCATATTATTACATTTGACTTTGAGTTAACTCAGACCTAGGTCTGGAA 458
 QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
 Db 459 TGCTGTCTTCTCTACTATTATCCAAAATATACATATCCACAGATCATTA 503
 RESULT 5
 ARI89269
 LOCUS ARI89269 2512 bp mRNA linear PRI 22-MAR-2001
 DEFINITION Homo sapiens UC28 protein (UC28) mRNA, complete cds.
 ACCESSION ARI89269
 VERSION ARI89269.1 GI:10441601
 KEYWORDS Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2512)
 AUTHORS An.G., Ng,A.Y., Meka,C.S., Luo,G., Bright,S.P., Cazares,L.,
 Wright,G.L. Jr. and Veltri,R.W.
 TITLE Cloning and characterization of UROC28, a novel gene overexpressed
 in prostate, breast, and bladder cancers
 JOURNAL Cancer Res. 60 (24), 7014-7020 (2000)
 MEDLINE 21028101
 PUBMED 11156405

REFERENCE 2 (bases 1 to 2512)
 AUTHORS An.G. and Veltri,R.W.
 TITLE Direct Submission
 JOURNAL Submitted (23-SEP-1999) UroSciences, UroCor, Inc., 840 Research
 Parkway, Oklahoma City, OK 73104, USA

FEATURES
 source
 1..2512
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6q23-q24"
 /tissue_type="prostate cancer"
 gene
 1..2512
 /gene="UC28"
 CDS
 99..506
 /gene="UC28"
 /codon_start=1
 /product="UC28 protein"
 /protein_id="AAG17117.1"
 /db_xref="GI:1041602"
 /translation="MRAFLRNOKYEDMHNIIHLQIRKLHRLSNFPLPGLAPETV
 LIPFCYKVRKKEKVKRSOKAFIDYSEOSHALLTPLOTHLMKSGSMCKSLSS
 EAILFTLTLQLTQLGLECLLYLSKTHPQII"
 polyA_signal
 2487..2492
 /gene="UC28"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,686-67 Length: 2512
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-09-974-546C-86 (1-135) x AF189269 (1-2512)

QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
 |||||
 Db 99 ATGAGGCGCTTCTTAAGGAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
 |||||

QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 |||||
 Db 159 CAGATCAGAAAAATTGAGGCACAGATTAACTTCCCAAGGCTACAGGCATTCTAGCT 218
 |||||

QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
 |||||
 Db 219 CCAGAAACTGTGCTCTTACCATCTGCTCAAGAGTATTTCGAAAAAAGAAAAAGTAAAA 278
 |||||

QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 |||||
 Db 279 AGAAGTCAAAAGGCAACAGATTCAATGATTATTCATAGAACAGTCACACCATCAATT 338
 |||||

QY 81 LeuThrProLeuGlnThrHisIleuThrMetLysGlySerSerMetLysCysSerSerLeu 100
 |||||
 Db 339 CTCACACCCCTTCGACACACACTTGACCATGAAGGTTCCTCAATGAAATGTTCTCATTTA 398
 |||||

QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 |||||
 Db 399 TCTTCGAAAGCCATATTATTCATGACTTTTGCAGTTAACTCAGACCCCTAGTCTCGAA 458
 |||||

QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
 |||||
 Db 459 TGCTGTCTTCTTACTTATCCAAACTATACATCCACAGATCATTA 503
 |||||

RESULT 6
 HS171N11
 LOCUS
 DEFINITION Human DNA sequence from clone RP1-171N11 on chromosome 6q23,
 complete sequence.
 ACCESSION AL031433.4 GI:4826442
 VERSION AL031433.4
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 129227)
 AUTHORS Williams,S.
 TITLE Direct Submission
 JOURNAL Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On May 13, 1999 this sequence version replaced gi:4582115.
 COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr6
 RP1-171N11 is from the library RP1-1 constructed by the group of
 Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pCYPAC2.

FEATURES
 Location/Qualifiers
 1..129227
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="q23"
 /clone="RP1-171N11"
 /clone_lib="RP1-1"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,196-65 Length: 129227
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0


```
DB: 9 Gaps: 0
US-09-974-546C-86 (1-135) x HSL17IN11 (1-129227)

QY 1 MetArgAlaPheLeuArgAenGlnLysTyrGluAspMetHisAsnIleHileLeu 20
Db 29892 ATGAGGGCCCTTTAAGGAACCAAGAAATATGAGGATATGCACATATATTACATTTTA 29951
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 29952 CAGATCAGAAAATTGAGGCACAGATTAACTTCCCAAGGCTACCGGCATTTCTAGCT 30011
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
Db 30012 CCAGAAACTGTGCTCTTACCATCTCTGCTACAGAGTATTTGCAAAAAGAAAGTAA 30071
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 30072 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 30131
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
Db 30132 CTCACACCCCTGCGACACACATTCGACCATGAAGGTTCTCAATGAATGTTCTCTCATTA 30191
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 30192 TCTTCAGAGCCCATATATTACATTCAGTTCCTGAGTTTACATCCAGACCCCTAGGCTCGAA 30251
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 30252 TGCTGTCTTCTACTATTATCCAAACTATATACATCCACAGATCATA 30296

RESULT 7
AC036236 186698 bp DNA linear HTG 01-SEP-2000
LOCUS Homo sapiens chromosome 6 clone RP11-133015, WORKING DRAFT
DEFINITION AC036236
ACCESSION AC036236
VERSION AC036236.2 GI:9958166
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 186698)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 186698)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (07-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Sep 1, 2000 this sequence version replaced gi:7523932.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu/gsc/index.shtml
Center project name: H_NH0133015
----- Project Information -----
Sequencing vector: p13; 100%
Chemistry: Dye-primer ET; 100% of reads
Assembly: Dye-terminator Big Dye; 0% of reads
Consensus quality: 183373 bases at least Q30
Consensus quality: 184432 bases at least Q40
Consensus quality: 185064 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 186098; sum-of-contigs
```

Quality coverage: 5.84 in Q20 bases; agarose-fp
Quality coverage: 5.57 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 4769: contig of 4769 bp in length
* 4770 4869: gap of unknown length
* 4870 11657: contig of 6788 bp in length
* 11658 11757: gap of unknown length
* 11758 28189: contig of 16432 bp in length
* 28190 28289: gap of unknown length
* 28290 53441: contig of 25052 bp in length
* 53442 102327: gap of unknown length
* 53442 102327: contig of 48886 bp in length
* 102328 102427: gap of unknown length
* 102428 182781: contig of 80354 bp in length
* 182782 182881: gap of unknown length
* 182882 186698: contig of 3817 bp in length.

FEATURES

source

1..186698
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-133015"

misc_feature

1..4769
/note="assembly_name:Contig10"

misc_feature

4870..11657
/note="assembly_name:Contig11"

misc_feature

11758..28189
/note="assembly_name:Contig12"

misc_feature

28290..53341
/note="assembly_name:Contig13"

misc_feature

53442..102327
/note="assembly_name:Contig14"

misc_feature

102428..182781
/note="assembly_name:Contig15
vector_side:right"
182882..186698
/note="assembly_name:Contig9"

ORIGIN

Alignment Scores:
Pred. No.: 3.3e-65 Length: 186698
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-974-546C-86 (1-135) x AC036236 (1-186698)

QY 1 MetArgAlaPheLeuArgAenGlnLysTyrGluAspMetHisAsnIleHileLeu 20
Db 159252 ATGAGGGCCCTTTAAGGAACCAAGAAATATGAGGATATGCACATATATTACATTTTA 159311
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159312 CAGATCAGAAAATTGAGGCACAGATTAACTTCCCAAGGCTACCGGCATTTCTAGCT 159371
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
Db 159372 CCAGAAACTGTGCTCTTACCATCTCTGCTACAGAGTATTTGCAAAAAGAAAGTAA 159431
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80

Db 159432 AGAAGTCAAAGGCCAACAGAGTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 159491

QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
|||||

Db 159492 CTCACACCCCTTCAGACACACTTGACCATGAAGGTTCTCATGAAATGTTCCCTCATTA 159551
|||||

QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
|||||

Db 159552 TCTTCAGAGCCCATATTATTCATTGACTTTTCAGTAACTCAGACCCCTAGTCTGGAA 159611
|||||

QY 121 CysCysLeuLeuTyrlleuSerLysThrIleHisProGlnIleIle 135
|||||

Db 159612 TGCTGTCTTCTACTTATCCAAACTATACATCCACACATCAT 159656
|||||

RESULT 8
AC096345/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-18J15, WORKING DRAFT SEQUENCE, 2
unordered pieces.
ACCESSION AC096345
VERSION AC096345.6 GI:30521870
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 259920)
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J. J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P.,
Biswal, K., Blair, J. J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, K., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgieff, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowls, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuewa, L., Louisegh, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindarcne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Nwaokeme, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willison, R., Wiczzyk, R., Wooden, H., Worley, K.,

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 259920)
Worley, K. C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 259920)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GEYU
Center clone name: CH230-18J15
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 240867 bases at least Q40
Consensus quality: 244203 bases at least Q30
Consensus quality: 245971 bases at least Q20
Estimated insert size: 252807; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 258395: contig of 258395 bp in length
* 258396 258495: Gap of unknown length
* 258496 259920: contig of 1425 bp in length.
FEATURES
Location/Qualifiers
1..259920
/organism="Rattus norvegicus"
/db_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-18J15"
1..1739
/note="wgs_end_extension
clone_end:T7"
8054_8862
/note="clone_boundary
clone_end:T7"

```
misc_feature
    site: EcoRI
    end_sequence: BH359975"
    complement (248170..249015)
    /note="clone boundary"
    clone_end: Sp6

misc_feature
    site: EcoRI
    end_sequence: BH360032"
    249451..251486
    /note="wgs_end_extension"
    clone_end: Sp6"
    256461..258395
    /note="wgs_end_extension"
    clone_end: Sp6"

ORIGIN
Alignment Scores:
Pred. No.: 98.4 Length: 259920
Score: 99.50 Matches: 38
Percent Similarity: 40.54 Conservative: 22
Best Local Similarity: 25.68% Mismatches: 41
Query Match: 14.34% Indels: 47
DB: 2 Gaps: 8

US-09-974-546C-86 (1-135) x AC096345 (1-259920)
QY 10 TyTGluAspMetHisAsnIleHleLeuGlnIleArgLysLeuArgHisArgLeu 29
Db 195470 TTGAGGATGTGAGAGGACGATATCTTTTCACACTAAGGTAGAAAGACATCATCG 195411
QY 30 SerAsnProArgLeuProGlyIle-----LeuAlaProGluThrValLeu 45
Db 195410 TCCACATCCACATTTTCCAGAGTAGGAGACTGACATTTCCGCCAAGAGAAATTATT 195351
QY 46 -----LeuProPheCysTyr----- 50
Db 195350 TCCAAATTCAGTTTCTCTCCCTTTCCTGTATCCAGCTCACTAATTTCCACATATTTCTTCT 195291
QY 51 -----LysValPheArgLysLys-----GluLysValLys 60
Db 195290 TTCTGCTCTTCTGTTTCAAAATCTCTTTTGGAAAGAACTGAGTAGTGAAATAAATATCA 195231
QY 61 ArgSerGln-----LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHis 78
Db 195230 AGAGCCAGCTGTGGGCAACAGAGATCACTAGCTACTGACCCCAATATCTCCTCT 195171
QY 79 AlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSer 98
Db 195170 TTCTCTCTTACACCACT-----ATTTCATGAAGTCTACA-----GTATCT 195129
QY 99 SerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGly 118
Db 195128 GCTGGTTCAAGCTCTGCACCT-----ACGTACAAT 195099
QY 119 LeuGluCysCysLeuLeuTyrLeu 126
Db 195098 TTACAGAAGTGTAGCATATCTT 195075

RESULT 9
AF291597
LOCUS Homo sapiens isolate Japanese 18 22q11.2 noncoding genomic
DEFINITION sequence.
ACCESSION AF291597
VERSION AF291597.1 GI:10953199
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 9833)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Worldwide DNA sequence variation in a 10-kilobase noncoding region
```

```
on human chromosome 22
Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE 20481912
PUBMED 11005839
REFERENCE 2 (bases 1 to 9833)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas
at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
77030, USA
FEATURES
source
1..9833
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="Japanese 18"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
misc_feature 1..9833
/note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.: 8.07 Length: 9833
Score: 95.00 Matches: 42
Percent Similarity: 34.10% Conservative: 17
Best Local Similarity: 24.28% Mismatches: 38
Query Match: 13.69% Indels: 76
DB: 9 Gaps: 5

US-09-974-546C-86 (1-135) x AF291597 (1-9833)
QY 4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHleHisIleLeuGlnLeuArg 23
Db 2955 TTTATAATTTCTCAAAATCTCTATGAGCAGRTATGAAATCTATATTTTCAGATACAG 3014
QY 24 LysLeuArgHisArgLysSerAsnProArgLeuProGlyIleLeuAlaProGluThr 43
Db 3015 AAACGTAGCTATAGGCTAAGTAATTTCCCTAGGCTT----- 3050
QY 44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
Db 3051 GTAGATCTGGGATTT----- 3065
QY 64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
Db 3066 ---GAAACCGGTTCTTAGACACTAAG-----GCTAGTCT 3098
QY 84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
Db 3099 TTAACATACACCTTTTACTGCTCCAAATAATTATGCTGAATGCTGTATCTGTCTGTCAGGA 3158
QY 104 AlaIle-----LeuPheThrLeu 109
Db 3159 GATATTTTCTTAATAAATAGGATGGAGGTGTGTGTTGGGGGAGCTTTCTAGCTTG 3218
QY 110 ThrLeuGlnLeuThrGlnThrLeuGly----- 118
Db 3219 TCATTCCTTATGGGCCACGAGTTGTGAGGATTAGACCAAAATCTTTGGGGAAGAGA 3278
QY 119 -----Leu 119
Db 3279 AAAATGGAACCATGGTCTGTCTCTGAATTTCTCCACAGATCCCTTTGGTCCCAACCTC 3338
QY 120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
Db 3339 CAAGTGAGGCTCATTTTATTTATCTCTCTAGCTCTACACCT 3377
RESULT 10
AF291608
LOCUS Homo sapiens isolate Berg 26 22q11.2 noncoding genomic sequence.
DEFINITION
```

ACCESSION AF291608
VERSION AF291608.1 GI:10953210
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 9432)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Pathy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Worldwide DNA sequence variation in a 10-kilobase noncoding region
on human chromosome 22
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE 20481912
PUBMED 11005839
REFERENCE 2 (bases 1 to 9432)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Pathy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas
at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
77030, USA
FEATURES
source
1..9432
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="Berg 26"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
misc_feature
1..9432
/note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.: 9.95 Length: 9432
Score: 94.00 Matches: 42
Percent Similarity: 34.10% Conservative: 17
Best Local Similarity: 24.28% Mismatches: 38
Query Match: 13.54% Indels: 76
DB: 9 Gaps: 5
US-09-974-546C-86 (1-135) x AF291608 (1-9432)
QY 4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
Db 2955 TTTATAATTTCTCAAAAATCCTATGAGCGAGGTATGTAATCTATATTTTGCAGATACAG 3014
QY 24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
Db 3015 AAACCTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT----- 3050
QY 44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
Db 3051 GTAGATCTGGGATTT----- 3065
QY 64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
Db 3066 ---GAAACCGGGTCTTAGACACTAAG-----CCTAGTCTCT 3098
QY 84 LeuGlnThrHisLeuThrMetLysGlySerMetLysCysSerSerLeuSerSerGlu 103
Db 3099 TTAACCTACACCCCTTACTGCCTCCAAAATATTATGCTGATGATCTCTGATATCTGTCAGGA 3158
QY 104 AlaIle-----LeuPheThrLeu 109
Db 3159 GATATTTTCTCTATAATAAATAGGATGGAGGTGTGTGTGGGGAGGCTTTCTAGCTTG 3218
QY 110 ThrLeuGlnLeuThrGlnThrLeuGly----- 118
Db 3219 TCATTCCTTATGGCCAGCAGGCTTGGTGAGGATTAGACCAAAATCTTTGGGGAAGAGA 3278
QY 119 -----Leu 119

Db 3279 AAAATGGAACCATGCTGTTCTGAATTTCTCCACAGATCCCTTTGGTCCCAACCTC 3338
QY 120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
Db 3339 CAAGTGAGGCTCATTTTATTCTCTCTAGCTCTACACCT 3377
RESULT 11
LOCUS AF291631
DEFINITION Homo sapiens isolate South African MeJe 22q11.2 noncoding genomic
sequence.
ACCESSION AF291631
VERSION AF291631.1 GI:10953233
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 9817)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Pathy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Worldwide DNA sequence variation in a 10-kilobase noncoding region
on human chromosome 22
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
MEDLINE 20481912
PUBMED 11005839
REFERENCE 2 (bases 1 to 9817)
AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
Pamilo, P., Trexler, M., Pathy, L., Jorde, L.B., Yu, N. and Li, W.-H.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas
at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
77030, USA
FEATURES
source
1..9817
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="South African MeJe"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
misc_feature
1..9817
/note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.: 10.4 Length: 9817
Score: 94.00 Matches: 42
Percent Similarity: 34.10% Conservative: 17
Best Local Similarity: 24.28% Mismatches: 38
Query Match: 13.54% Indels: 76
DB: 9 Gaps: 5
US-09-974-546C-86 (1-135) x AF291631 (1-9817)
QY 4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
Db 2939 TTTATAATTTCTCAAAAATCCTATGAGCGAGGTATGTAATCTATATTTTGCAGATACAG 2998
QY 24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
Db 2999 AAACCTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT----- 3034
QY 44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
Db 3035 GTAGATCTGGGATTT----- 3049
QY 64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
Db 3050 ---GAAACCGGGTCTTAGACACTAAG-----CCTAGTCTCT 3082
QY 84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103

```

Db      3083 TTAACGTACCCCTTACTGCCTCCAAAAATATGCTGAATGCTCTGATCTCTGTCAGGA 3142
QY      104 AlaIle-----:::|||||:::|||||:::|||||:::|||||:::|||||
Db      3143 GATATTTTCTCTAATAAAATAGGATGGAGGTGTGTGTGGGGGAGCTTTCTAGCTTG 3202
QY      110 ThrLeuGlnLeuThrGlnThrLeuGly-----118
Db      3203 TCATTCCTTATGGCCAGCAGGTGTGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3262
QY      119 -----Leu 119
Db      3263 AAAATGGAACCATGCTGCTGCTCCTGAATTCCTCAACAGATCCCTTGGTCCCAACCTC 3322
QY      120 GluCysCysLeuLeuTyrLeuSerTyrThrLeuHisPro 132
Db      3323 CAAGTGAGGCTCATTTTATCTCTCTAGCTCTACACCT 3361

RESULT 12
AF291632      9817 bp DNA linear PRI 24-OCT-2000
LOCUS      Homo sapiens isolate South African Kung 22q11.2 noncoding genomic
DEFINITION
ACCESSION      AF291632
VERSION
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 9817)
Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
Worldwide DNA sequence variation in a 10-kilobase noncoding region
on human chromosome 22
Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
JOURNAL      20481912
MEDLINE      11005839
REFERENCE      2 (bases 1 to 9817)
AUTHORS      Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
Direct Submission
TITLE      Submitted (01-AUG-2000) Human Genetics Center, University of Texas
JOURNAL      at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
77030, USA
FEATURES
source      1..9817
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="South African Kung"
/db_xref="taxon:9606"
/chromosomes="22"
/map="22q11.2"
misc_feature      1..9817
/note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.:      10.4      Length:      9817
Score:      94.00      Matches:      42
Percent Similarity:      34.10%      Conservative:      17
Best local Similarity:      24.28%      Mismatches:      38
Query Match:      13.54%      Indels:      76
DBs:      9      Gaps:      5

US-09-974-546C-86 (1-135) x AF291632 (1-9817)

QY      4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
Db      2939 TTTATAATTTCTCAAAATCCTATGAGGCAGGATGTAATCTATATTTTGCAGATACAG 2998
QY      24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43

```

```

Db      2999 AAACGTAGTCATAGCTAAGTAATTTCCCTAGGCTT-----3034
QY      44 ValLeuLeuProPheCysTyrIysValPheArgTylsLysGluLysVallysArgSerGln 63
Db      3035 GTAGATCTGGGATTT-----3049
QY      64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisHisAlaIleLeuThrPro 83
Db      3050 ---GAACCGGGTTCTTAGACATAAG-----GCTAGTCCT 3082
QY      84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
Db      3083 TTAACACACCCCTTACTGCCTCCAAAAATATGCTGAATGCTCTGATATCTGTGACGA 3142
QY      104 AlaIle-----LeuPheThrLeu 109
Db      3143 GATATTTTCTCTAATAAAATAGGATGGAGGTGTGTGGGGGAGCTTTCTAGCTTG 3202
QY      110 ThrLeuGlnLeuThrGlnThrLeuGly-----118
Db      3203 TCATTCCTTATGGCCAGCAGGTGTGTGAGGATTAGAGCCAAAAATCTTTGGGGAAGAGA 3262
QY      119 -----Leu 119
Db      3263 AAAATGGAACCATGCTGCTGCTCCTGAATTCCTCAACAGATCCCTTGGTCCCAACCTC 3322
QY      120 GluCysCysLeuLeuTyrLeuSerTyrThrLeuHisPro 132
Db      3323 CAAGTGAGGCTCATTTTATCTCTCTAGCTCTACACCT 3361

RESULT 13
AF291644      9830 bp DNA linear PRI 24-OCT-2000
LOCUS      Homo sapiens isolate Nigerian 9 22q11.2 noncoding genomic sequence.
DEFINITION
ACCESSION      AF291644
VERSION      AF291644.1 GI:10953246
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 9830)
Zhao,Z., Jin,L., Fu,Y.-X., Ramsay,M., Jenkins,T., Leskinen,E.,
Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
Worldwide DNA sequence variation in a 10-kilobase noncoding region
on human chromosome 22
Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)
JOURNAL      20481912
MEDLINE      11005839
REFERENCE      2 (bases 1 to 9830)
AUTHORS      Pamilo,P., Trexler,M., Patthy,L., Jorde,L.B., Yu,N. and Li,W.-H.
Direct Submission
TITLE      Submitted (01-AUG-2000) Human Genetics Center, University of Texas
JOURNAL      at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
77030, USA
FEATURES
source      1..9830
/organism="Homo sapiens"
/mol_type="genomic DNA"
/isolate="Nigerian 9"
/db_xref="taxon:9606"
/chromosomes="22"
/map="22q11.2"
misc_feature      1..9830
/note="noncoding region"
ORIGIN
Alignment Scores:
Pred. No.:      10.4      Length:      9830
Score:      94.00      Matches:      42
Percent Similarity:      34.10%      Conservative:      17

```

Best Local Similarity:	24.28%	Mismatches:	38
Query Match:	13.54%	Indels:	76
DB:	9	Gaps:	5

US-09-974-546C-86 (1-135) x AF291644 (1-9830)

QY	4	PheLeuArgAsnGlnIlystyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg	23
Db	2955	TTTTATATTTCTCAAAAATCCTTAGAGCCAGGTATGTAATCTATATTTTTGCGATACAG	3014
QY	24	LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr	43
Db	3015	AAACTGAGTCATAGGCTAAGTAATTTCCCTAGGCTT	3050
QY	44	ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln	63
Db	3051	GTAGATCTGGGATTT	3065
QY	64	LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIleLeuThrPro	83
Db	3066	---GAAACCGGTTCTAGACACTAAG---	3098
QY	84	LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu	103
Db	3099	TTAACTACACCCCTTTACTGCCTCCAAAAATTATGCTGAATCTGTGATATCTGTCTCAGGA	3158
QY	104	AlaIle---	109
Db	3159	GATATTTTCTTAATAAATAGGATGGAGGTGTGTGTGGGGAGCTTTCTAGCTTG	3218
QY	110	ThrLeuGlnLeuThrGlnThrLeuGly---	118
Db	3219	TCATTCCTTATGGGCCAGCAGGTGTGTGAGGATTAGACCAGAAAAATCTTTGGGGAGAGAGA	3278
QY	119	---	119
Db	3279	AAATGGAACCATGGTCTGTCTCTCTGAATTCTCCAAACAGATCCCTTTGGTCCCAACCTC	3338
QY	120	GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro	132
Db	3339	CAAGTAGGCTCATATTTATCTCTCTAGCTCTACACCT	3377

RESULT	14
LOCUS	AF291587
DEFINITION	Homo sapiens isolate Chinese 1 22q11.2 noncoding genomic sequence.
ACCESSION	AF291587
VERSION	AF291587.1 GI:10953189
KEYWORDS	.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
Eukaryota;	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia;	Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 9833)
AUTHORS	Zhao, Z., Jin, L., Fu, Y.-X., Ramey, M., Jenkins, T., Leskinen, E., Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H. Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22
TITLE	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)

PURMED	11005819	
REFERENCE	2 (bases 1 to 9833)	
AUTHORS	Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E., Pamilo, P., Tretliar, M., Patthy, L., Jorde, L. B., Yu, N. and Li, W.-H.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-AUG-2000) Human Genetics Center, University of Texas at Houston Health Science Center, 6901 Bertner Ave., Houston, TX 77030. USA	

```

FEATURES
    source
    Location/Qualifiers
1. .9833
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /isolate="Chinese 1"

```

```

/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
1..9833
/note="noncoding region"

```

ORIGIN

Alignment Scores:		
Pred. No.:	10.4	Length: 9833
Score:	94.00	Matches: 42
Percent Similarity:	34.10%	Conservative: 17
Best Local Similarity:	24.28%	Mismatches: 38
Query Match:	13.54%	Indels: 76
DB:	9	Gaps: 5

US-09-974-546C-86 (1-135) x AF291587 (1-9833)

Qy	4	PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleHisIleLeuGlnIleArg	23
Db	2955	TTTATAATTCTCAAAAACTCTATGAGGAGGATGTAATCTATATTTTGCAGATACAG	3014
Qy	24	LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr	43
Db	3015	AAACTGAGTCATAGGCTAGTAATTTCCCTAGGCTT	3050
Qy	44	ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln	63
Db	3051	GTAGATCTGGGATT	3065
Qy	64	LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIleLeuThrPro	83
Db	3066	---GAAACCGGGTCTTAGACTAG---	3098
Qy	84	LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeuSerSerGlu	103
Db	3099	TTAACTACACCCCTTTACTGCCTCCAAAAATTATGCTGAATGCTCTGATATCTTGTGAGGA	3158
Qy	104	AlaIle---	109
Db	3159	GATATTTTCTCTAATAAAATAGGATGGAGGTGTGTGTGGGGGAGCTTCTTAGCTTG	3218
Qy	110	ThrLeuGlnLeuThrGlnThrLeuGly---	118
Db	3219	TCATTCCTTATGGCCAGCAGGTTGTTGAGGATTAGAGCCAAAAATCTTTGGGGGAAGAGA	3278
Qy	119	---	119
Db	3279	AAATGGAAACCATGGTCTGTCTCTGAATTTCTCAACAGATCCCTTTGGTCCCAACCTC	3338
Qy	120	GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro	132
Db	3339	CAAGTAGGCTCATTTTATTTCTCTAGCTCTACACCT	3377

RESULT 15

AF291588	AF291588	9833 bp	DNA	linear	PRI 24-OCT-2000
LOCUS					
DEFINITION	Homo sapiens isolate Chinese 2	22q11.2	noncoding	genomic	sequence.
ACCESSION	AF291588				
VERSION	AF291588.1	GI:10953190			

SOURCE Homo sapiens (human)

ORGANISM

taxon: Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

Authors Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E., Pamiilo, P., Trexler, M., Pathy, L., Jorde, L.B., Yu, N. and Li, W.-H.

Title Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22

JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (21), 11354-11358 (2000)

MEDLINE 20481912

PUBMED
11005839

REFERENCE 2 (bases 1 to 9833)

1000

AUTHORS Zhao, Z., Jin, L., Fu, Y.-X., Ramsay, M., Jenkins, T., Leskinen, E.,
 Pamilo, P., Trexler, M., Patthy, L., Jorde, L.B., Yu, N. and Li, W.-H.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2000) Human Genetics Center, University of Texas
 at Houston Health Science Center, 6901 Bertner Ave., Houston, TX
 77030, USA

FEATURES Location/Qualifiers
 source 1..9833
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /isolate="Chinese 2"
 /db_xref="taxon:9606"
 /chromosome="22"
 /map="22q11.2"
 misc_feature 1..9833
 /note="noncoding region"

ORIGIN

Alignment Scores:
 Pred. No.: 10.4 Length: 9833
 Score: 94.00 Matches: 42
 Percent Similarity: 34.10% Conservative: 17
 Best Local Similarity: 24.28% Mismatches: 38
 Query Match: 13.54% Indels: 76
 DB: 9 Gaps: 5

US-09-974-546C-86 (1-135) x AF291588 (1-9833)

QY 4 PheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeuGlnIleArg 23
 Db 2955 TTTATAATTCTCATAAATCTATAGGCAGGTATGTAATCTATATTTTGCAGATACAG 3014
 QY 24 LysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAlaProGluThr 43
 Db 3015 AAACGTAGTCATAGGCTAAGTAATTCCTTAGGCTT----- 3050
 QY 44 ValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLysArgSerGln 63
 Db 3051 GTAGATCTGGGATT----- 3065
 QY 64 LysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIleLeuThrPro 83
 Db 3066 ---GAAACCGGTTCTTAGACACTAG-----GCTAGTCCT 3098
 QY 84 LeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeuSerSerGlu 103
 Db 3099 TTAACATACACCCCTTACTGCTCCAAAATATGCTGAATGCTCTGATATCCTGTCCAGGA 3158
 QY 104 AlaIle-----LeuPheThrLeu 109
 Db 3159 GATATTTTCTTAATAAATAGGATGGAGGTGTGTGTGGGGAGCTTTCTAGCTTG 3218
 QY 110 ThrLeuGlnLeuThrGlnThrLeuGly----- 118
 Db 3219 TCATTCTTATGGCCAGCAGGTGTGTGAGGATTAGACCAAAATCTTTGGGGAAGAGA 3278
 QY 119 -----Leu 119
 Db 3279 AAAATGGAACCATGGTCTGTCTCTGAATTCCTCAACAGATCCCTTTGGTCCCAACCTC 3338
 QY 120 GluCysCysLeuLeuTyrLeuSerLysThrIleHisPro 132
 Db 3339 CAAGTAGGCTCATTTTATTCTCTCTAGCTCTACACCT 3377

Search completed: September 9, 2005, 14:45:55
 Job time : 2499 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:02:29 ; Search time 309 Seconds

(without alignments)
2586.295 Million cell updates/sec

Title: US-09-974-546C-86

Perfect score: 694

Sequence: 1 MRAFLRNQYEDMNIHL.....TLGLECCLLYLSTKTHPQII 135

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO spool/US09974546/runat_07092005_174501_20885/app_query.fasta_1.654
-DB=N_Geneseq_16Dec04 -QFMT=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosun62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CGN 1.1.468 @runat_07092005_174501_20885 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	2087	3	Aas04000 Biomarker
2	694	100.0	2088	4	Aaz87583 Prostate
3	694	100.0	2505	4	Aas04001 Biomarker
4	694	100.0	2506	3	Aaz87584 Prostate
5	90	13.0	1113	13	Adt45647 Bacterial

6	90	13.0	1146	13	ADS46591	Adt46591 Bacterial
7	89.5	12.9	4019	4	ABL27798	Abt27798 Drosophil
c 8	83.5	12.0	532	3	AAC38626	Aac38626 Arabidops
c 9	83	12.0	146547	8	ABZ80817	Abz80817 Human pho
c 10	82	11.8	24601	2	AAX13160	Aax13160 Enterococ
c 11	82	11.8	24601	6	ABS98955	Abs98955 Enterococ
c 12	81.5	11.7	133462	13	ABD32622	Abd32622 Mouse can
c 13	81	11.7	381	8	ABX44573	Abx44573 Bovine ES
c 14	81	11.7	110000	12	ADN46845_11	Continuation (12 o
c 15	81	11.7	110000	12	ADN47591_09	Continuation (10 o
c 16	81	11.7	110000	12	ADN46123_11	Continuation (12 o
c 17	81	11.7	110000	12	ADN47209_09	Continuation (10 o
c 18	81	11.7	110000	12	ADN46464_11	Continuation (12 o
c 19	81	11.7	110000	12	ADN47960_09	Continuation (10 o
c 20	80.5	11.6	1093	2	AAT95761	Aac95761 Arabidops
c 21	80.5	11.6	1093	3	AAC65293	Aac65293 Arabidops
c 22	80	11.5	1693	12	ADQ24476	Adq24476 Human sof
c 23	79.5	11.5	262	7	ADS31066	Adt31066 Human gen
c 24	79.5	11.5	585	5	ABV54205	Abv54205 Human pro
c 25	79.5	11.5	1464	3	AAC47238	Aac47238 Arabidops
c 26	79.5	11.5	6196	2	AAV13168	Aav13168 Complete
c 27	79.5	11.5	6243	2	AAV13841	Aav13841 Complete
c 28	79.5	11.5	6503	2	AAV13169	Aav13169 Complete
c 29	79.5	11.5	7379	2	AAV13176	Aav13176 Complete
c 30	79.5	11.5	8618	2	AAV18741	Aav18741 Complete
c 31	79.5	11.5	8792	2	AAV18745	Aav18745 Complete
c 32	78.5	11.3	386	8	ABX45278	Abx45278 Bovine ES
c 33	78.5	11.3	398	4	AAI89039	Aai89039 Human pol
c 34	78.5	11.3	4065	12	ADQ65017	Adq65017 Novel hum
c 35	78.5	11.3	26496	5	ABA16966	Abal6966 Human ner
c 36	78	11.2	3328	4	ABL18311	Abt18311 Drosophil
c 37	77.5	11.2	389	8	ABX39152	Abx39152 Bovine ES
c 38	77.5	11.2	567	10	ABX56974	Abx56974 Arabidops
c 39	77.5	11.2	2411	11	ADMO2309	Adm02309 Human cDN
c 40	77.5	11.2	2437	12	ADP22463	Adp22463 Sea-squir
c 41	77.5	11.2	2661	4	ABL26351	Abt26351 Drosophil
c 42	77.5	11.2	3435	4	ABL02615	Abt02615 Drosophil
c 43	77.5	11.2	3958	4	ABL07870	Abt07870 Drosophil
c 44	77.5	11.2	4122	4	ABL07868	Abt07868 Drosophil
c 45	77.5	11.2	6010	10	ADF50278	Adf50278 Human PFM

ALIGNMENTS

RESULT 1
AAS04000
ID AAS04000 standard; cDNA; 2087 BP.
XX
AC AAS04000;
XX
DT 29-AUG-2001 (first entry)
XX
DE Biomarker UC band 28 #2, used in diagnosis and prognosis of cancer.
KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 99..506
FT /*tag= a
FT /product= "Prostate cancer marker protein"
XX
PN US6218529-B1.
XX
PD 17-APR-2001.
XX
PF 12-JUN-1998; 98US-00097199.
PR 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.

XX (UROC-) UROCOR INC.
XX An G, O'hara SM, Ralph D, Veltri R;
XX WPI; 2001-289849/30.
XX P-PSDB; AAU02174.
XX New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX Claim 2; Col 117-121; 78pp; English.
XX The sequence represents nucleic acid biomarker UC band 28 #2, used in
CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
CC sequences can be used as hybridisation probes and primers that
CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
CC nucleic acid markers can be used to produce antibodies for the detection
CC of prostate, breast or bladder cancer. The nucleic acids can be used as
CC targets for therapeutic intervention in these diseases, in the
CC identification and isolation of full-length gene sequences, including
CC regulatory elements for gene expression, from genomic human DNA
CC libraries, as hybridisation probes for screening genomic human DNA
CC libraries. The kits comprising the nucleic acid sequences are useful for
CC detecting bladder, breast or prostate cancer cells in a biological sample
XX Sequence 2087 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.6e-76 Length: 2087
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-974-546C-86 (1-135) x AAS04000 (1-2087)

QY 1 MetArgAlaPheLeuArgGlnIysTyrGluAspMetHisAsnIleHisIleLeu 20
Db 99 ATGAGGCGCTTCTTAAGGAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAATTTGAGGCACAGATTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysVallys 60
Db 219 CCAGAAACTGTGCTCTTACCATTTCTGCTACAAGGTATTTCGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisIleAlaIle 80
Db 279 AGAAGTCAAAAGGCAACAGAGTTCTATTGATTATTCATAGAACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
Db 339 CTCACACCTTGCAGACACACTTGACCATGAAGGTCTCTCAATGNAAGTTCCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 399 TCTTCAGAGCCATATTATTCATATGACTTTGCAAGTTAACTCAGACCCTAGGCTCGGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 459 TGCTGTCTTCTTACTTATCCAAAACATATACATCCACAGATCATATA 503

RESULT 2
ID AA287583
XX AA287583 standard; DNA; 2088 BP.
AC AA287583;

XX 19-APR-2000 (first entry)
XX Prostate disease marker UC Band #28.
XX Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
diagnosis; PCR primer; ss.
XX Homo sapiens.
XX WO9964631-A1.
XX 16-DEC-1999.
XX 11-JUN-1999; 99WO-US013151.
XX 12-JUN-1998; 98US-00097199.
XX (UROC-) UROCOR INC.
XX An G, O'hara SM, Ralph D, Veltri RW;
XX WPI; 2000-116557/10.
XX P-PSDB; AAY59295.
XX Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX Claim 1; Page 182-183; 191pp; English.
XX The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer. The
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences from various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX Sequence 2088 BP; 740 A; 406 C; 337 G; 604 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 2.6e-76 Length: 2088
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-974-546C-86 (1-135) x AA287583 (1-2088)

QY 1 MetArgAlaPheLeuArgAsnGlnIysTyrGluAspMetHisAsnIleHisIleLeu 20
Db 99 ATGAGGCGCTTCTTAAGGAACAGAAATATGAGGATATGCACAAATATTATTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAATTTGAGGCACAGATTAACTTCCCAAGGCTACCAGGCATTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysVallys 60
Db 219 CCAGAAACTGTGCTCTTACCATTTCTGCTACAAGGTATTTCGAAAAAAGAAAAAGTAAAA 278

QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 Db 279 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
 QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 Db 339 CTCACACCTTCGAGACACACTTGACCATGAAAGGTTCTCATGAAATGTTCCATTA 398
 QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 Db 399 TCTTCAGAAAGCCATATTATTACATTGACTTTGCAGTTAACTCAGACCCCTAGGCTCGAA 458
 QY 121 CysCysLeuLeuTyrIleuSerLysThrIleHisProGlnIleIle 135
 Db 459 TGCTGTCTTCTTACTTATCCAAACTATATACATCCACAGATCATATA 503
 RESULT 3
 AAS04001
 ID AAS04001 standard; cDNA; 2505 BP.
 XX AAS04001;
 AC AAS04001;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE Biomarker UC band 28 #3, used in diagnosis and prognosis of cancer.
 XX
 KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
 KW benign prostatic hyperplasia; BPH; therapeutic; human; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Location/Qualifiers
 FT 99..506
 FT /*tag= a
 FT /product= "Prostate cancer marker protein"
 XX
 XX
 PN US6218529-B1.
 XX
 PD 17-APR-2001.
 XX
 PF 12-JUN-1998; 98US-00097199.
 XX
 PR 31-JUL-1995; 95US-0001655P.
 PR 11-JAN-1996; 96US-0013611P.
 PR 31-JUL-1996; 96US-00692787.
 XX
 PA (UROC-) UROCOR INC.
 XX
 PI An G, O'hara SM, Ralph D, Veltri R;
 XX
 DR WPI; 2001-289849/30.
 DR P-PSDB; AAU02175.
 XX
 PT New nucleic acids as biomarkers and targets useful for detecting,
 PT diagnosing, prognosing, and in developing treatments for prostate, breast
 PT and bladder cancer.
 XX
 PS Claim 2; Col 121-125; 78pp; English.
 XX
 CC The sequence represents nucleic acid biomarker UC band 28 #3, used in
 CC detection of prostate, breast and bladder cancer. Biomarker nucleic acid
 CC sequences can be used as hybridisation probes and primers that
 CC specifically hybridise to prostate cancer, benign prostatic hyperplasia
 CC (BPH), bladder cancer or breast cancer markers. Proteins encoded by the
 CC nucleic acid markers can be used to produce antibodies for the detection
 CC of prostate, breast or bladder cancer. The nucleic acids can be used as
 CC targets for therapeutic intervention in these diseases, in the
 CC identification and isolation of full-length gene sequences, including
 CC regulatory elements for gene expression, from genomic human DNA
 CC libraries, as hybridisation probes for screening genomic human DNA
 CC libraries. The kits comprising the nucleic acid sequences are useful for
 CC detecting bladder, breast or prostate cancer cells in a biological sample

SQ Sequence 2505 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3.39e-76 Length: 2505
 Score: 694.00 Matches: 135
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0
 US-09-974-546C-86 (1-135) x AAS04001 (1-2505)
 QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
 Db 99 ATGAGGGCCCTTCTTAAGGAACCCAGAAATATGAGGATATGCACATATTATTTCACATTTTA 158
 QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
 Db 159 CAGATCAGAAAATTGAGGCACAGATTAAGTAACTTCCCAAGGCTACCGGCATTCTAGCT 218
 QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGlyLysValLys 60
 Db 219 CCAGAAACTGTGCTCTTACCATTTCTGCTACAGGTTATTCGAAAAAAGAAAAAGTAA 278
 QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
 Db 279 AGAAGTCAAAAGGCAACAGAGTTTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338
 QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
 Db 339 CTCACACCTTCGACACACACTTGACCATGAAAGGTTCTCAATGAAATGTTCCATTA 398
 QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
 Db 399 TCTTCAGAAAGCCATATTATTACATTGACTTTGCAGTTAACTCAGACCCCTAGGCTCGAA 458
 QY 121 CysCysLeuLeuTyrIleuSerLysThrIleHisProGlnIleIle 135
 Db 459 TGCTGTCTTCTTACTTATCCAAACTATATACATCCACAGATCATATA 503
 RESULT 4
 AAZ87584
 ID AAZ87584 standard; DNA; 2506 BP.
 XX
 AC AAZ87584;
 XX
 DT 19-APR-2000 (first entry)
 XX
 DE Prostate disease marker UC Band #28 splice variant.
 XX
 KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
 KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection;
 KW diagnosis; PCR primer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9964631-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 11-JUN-1999; 99WO-US013151.
 PR 12-JUN-1999; 98US-00097199.
 XX
 PA (UROC-) UROCOR INC.
 XX
 PI An G, O'hara SM, Ralph D, Veltri RW;
 XX
 DR WPI; 2000-116557/10.
 DR P-PSDB; AAY59296.
 XX
 PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
 PT breast and bladder cancer.

XX Claim 1; Page 184-186; 191pp; English.

XX The invention provides nucleic acid markers of prostate, breast and

CC bladder cancer. The markers are indicators of malignant transformation of

CC prostate, breast and bladder tissues and are diagnostic of the potential

CC for metastatic spread of malignant prostate tumours. The nucleic acid can

CC also be used as targets for therapeutic intervention in prostate cancer,

CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The

CC markers may be used to design specific probes and primers, for the rapid

CC analysis of prostate, bladder or breast biopsy samples. The probes and

CC primers may also be used for in situ hybridization or in situ PCR

CC detection and diagnosis. They may also be used to identify and isolate

CC full length gene sequences from various DNA libraries. Antibodies against

CC the polypeptide products of the markers can be used to treat prostate

CC cancer, bladder cancer or breast cancer. The encoded proteins may be used

CC to detect antibodies. The proteins and antibodies can be used in

CC immunodetection methods for detecting or quantifying the cancers, and for

CC clinical diagnosis of these cancers. The antibodies may also be used for

CC radioimaging to quantify and localize the encoded proteins

XX SQ Sequence 2506 BP; 822 A; 519 C; 446 G; 718 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	3.39e-76	Length:	2506
Score:	694.00	Matches:	135
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-09-974-546C-86 (1-135) x AAZ87584 (1-2506)

QY 1 MetArgAlaPheLeuArgAsnGlnIlyTyRGlUaPwMetHisAsnIleIleHisIleLeu 20

DB 99 ATGAGGCGCTCTTAAAGGAACAGAAATATAGGATATGCACAAATATATTACATTTTA 158

QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40

DB 159 CAGATCAGAAATTCGAGGCACAGATTAGTAATCTCCCAAGGCTACCAGCATTCAGCT 218

QY 41 ProGluThrValLeuLeuProPheCysTyRysValPheArgLysLysGluLysVallys 60

DB 219 CCAGAAACTGTCTCTTACCATCTTGCTACAAGTATTTCGAAAAAAGAAAAAGTAAA 278

QY 61 ArgSerGlnLysAlaThrGluPheIleAepTyRSerIleGluGlnSerHisAlaIle 80

DB 279 AGAAGTCAAAAGGCACAGATTCATTGATTATTCATAGAACAGTCACACCATGCAATT 338

QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100

DB 339 CTCACACCTTCGACACACACTTGACCATGAAGGTTCCTCATGNAATGTTCTCATTA 398

QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120

DB 399 TCTTCAGAGCCATATTATTACATTGACTTTGCGAGTTAACTCAGACCCCTAGGTCTGAA 458

QY 121 CysCysLeuLeuTyRLeuSerLysThrIleHisProGlnIleIle 135

DB 459 TGCTGTCTCTCTACTTATCCAAAATCTATACATCCACATCATTA 503

RESULT 5

ADT45647

ID ADT45647 standard; cDNA; 1113 BP.

XX AC ADT45647;

XX 02-DEC-2004 (first entry)

XX Bacterial polynucleotide #20398.

XX Recombinant DNA construct; transformed plant; improved plant property;

XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

KW pathogen tolerance; pest tolerance; plant disease resistance;

KW cell cycle pathway modification; plant growth regulator;

KW homologous recombination; seed oil yield; protein yield; carbohydrate;

KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

XX bacterial polynucleotide; gene; ss.

OS Bacteria.

XX US2003233675-A1.

PN 18-DEC-2003.

PD 20-FEB-2003; 2003US-00369493.

PF 21-FEB-2002; 2002US-0360039P.

PR (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.

PA (CHEN/) CHEN X.

PA (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

PI WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide

CC for expression of a polynucleotide encoding a polypeptide from a

CC microbial source, useful for producing plants with improved properties.

PS Claim 1; SEQ ID NO 44085; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a

CC promoter functional in a plant cell, where the promoter is positioned to

CC provide for expression of a polynucleotide encoding a polypeptide from a

CC microbial source. The invention also relates to a transformed plant

CC comprising the recombinant DNA construct and a method of producing a

CC transformed plant having an improved property. The plant is a crop plant

CC such as maize or soybean. The method of producing a transformed plant

CC having an improved property comprises transforming a plant with the

CC recombinant DNA construct and growing the transformed plant, where the

CC polynucleotide or polypeptide is useful for improving plant properties.

CC The recombinant DNA construct is useful for producing plants with

CC improved plant properties, e.g. improved cold, heat or drought tolerance,

CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,

CC increased resistance to plant disease, better growth rate by modification

CC of the cell cycle pathway with plant growth regulators, increased rate of

CC homologous recombination, modified seed oil or protein yield and/or

CC content, improved yield by modification of carbohydrate, nitrogen or

CC phosphorus use and/or uptake, by modification of photosynthesis or by

CC providing improved plant growth and development under at least one stress

CC condition, improved lignin production or improved galactomannan

CC production. This sequence represents a bacterial polynucleotide used in

CC the scope of the invention. Note: The sequence data for this patent did

CC not form part of the printed specification but was obtained in electronic

CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 1113 BP; 379 A; 204 C; 242 G; 288 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.53	Length:	1113
Score:	90.00	Matches:	34
Percent Similarity:	44.88%	Conservative:	23
Best Local Similarity:	26.77%	Mismatches:	38
Query Match:	12.97%	Indels:	32
DB:	13	Gaps:	7

US-09-974-546C-86 (1-135) x ADT45647 (1-1113)

QY 13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu----- 29

DB 61 ATGCACAACTTA-----GCCATTAAGCTGAGAGAGAGGCGCATGAAGTGGGAATAGTC 114

QY 30 ---SerAsnPhePro----- 33
 Db 115 ACAACACAGGCGCCAGGAAAGAGAGAGCTTAAGAGATATGAATAGAGCTCATA 174
 QY 34 ArgLeuProGlyLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
 Db 175 AAGATCCAGGAATTATAAGTCT-----TTTATAGATGTA 210
 QY 54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle---AspTyrSer 72
 Db 211 AATTTAACTCATGATGAATCTCAGAGAGCTCAACGAATTTCTTGAAGACTTCGAC 270
 QY 73 IleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
 Db 271 ATAATTCATTTCTCATATGCA---TTCACACCTCTCTCTTAAAGGCTTTAAAGCTGGA 327
 QY 93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
 Db 328 AAGAATATGGAAGGAGACCTGCTTAACAACACTCACAGCATTTCTTGGCCCATGAATCA 387
 QY 112 GlnLeuThrGlnThrLeuGly 118
 Db 388 AAGCTCTGGGATACCTTTGGGG 408
 RESULT 6
 ADS46591
 ID ADS46591 standard; cDNA; 1146 BP.
 XX
 AC ADS46591;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polynucleotide #1334.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.
 XX
 OS Bacteria.
 XX
 PN US2003233675-A1.
 XX
 XX 18-DEC-2003.
 XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 XX 21-FEB-2002; 2002US-0360039P.
 PR
 XX (CAOX/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX WPI; 2004-061375/06.
 XX
 XX New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX
 XX Claim 1; SEQ ID NO 25021; 122pp; English.
 PS
 XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a

CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 1146 BP; 388 A; 207 C; 253 G; 298 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.553 Length: 1146
 Score: 90.00 Matches: 34
 Percent Similarity: 44.88% Conservative: 23
 Best Local Similarity: 26.77% Mismatches: 38
 Query Match: 12.97% Indels: 32
 DB: 13 Gaps: 7
 US-09-974-546C-86 (1-135) x ADS46591 (1-1146)
 QY 13 MetHisAsnIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu----- 29
 Db 70 ATGCACAACTTA-----GCCATAAAGCTGAGAGAAAGAGCGCATGAAGTGGGAATAGTC 123
 QY 30 ---SerAsnPhePro----- 33
 Db 124 ACAACAACAGCGCCCGGAAAGAGAGCTTAAGAGATATGAATAGAGCTCATA 183
 QY 34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
 Db 184 AAGATCCAGGAATTATAAGTCT-----TTTATAGATGTA 219
 QY 54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle---AspTyrSer 72
 Db 220 AATTTAACTCATGATGAATCTCAGAGAGCTCAACGAATTTCTTGAAGACTTCGAC 279
 QY 73 IleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
 Db 280 ATAATTCATTTCTCATCATGCA---TTCACACCTCTCTCTTAAAGGCTTTAAAGCTGGA 336
 QY 93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
 Db 337 AAGAATATGGAAGGAGACCTGCTTAACAACACTCACAGCATTTCTTGGCCCATGAATCA 396
 QY 112 GlnLeuThrGlnThrLeuGly 118
 Db 397 AAGCTCTGGGATACCTTTGGGG 417
 RESULT 7
 ABL27798
 ID ABL27798 standard; DNA; 4019 BP.
 XX
 AC ABL27798;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 34867.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

XX Claim 1; SEQ ID NO 34867; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 4019 BP; 1158 A; 966 C; 927 G; 968 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.93 Length: 4019
Score: 89.50 Matches: 35
Percent Similarity: 39.84% Conservative: 16
Best Local Similarity: 27.34% Mismatches: 40
Query Match: 12.90% Indels: 37
DB: 4 Gaps: 5

US-09-974-546C-86 (1-135) x ABL27798 (1-4019)

QY 18 HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGly 37
DB 1389 CATAGACTAAAGCAGCAAGGTGCAACCGTAATTATTAATTTTACAGCTGTTCCAGCG 1448
QY 38 IleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGlu 57
DB 1449 CATAGGCAGCCCATCA-----TTCACATCGCATCAAGACGAAAGAAAGCA 1499
QY 58 LysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluInSerHis 77
DB 1500 CGTCAGAGCAGCGAAACTAAGTGTACTTCCAAAGATA-----CAT 1538
QY 78 HisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCys 97
DB 1539 TGTACAAATTTACATGCCGCTG-----AGGTGC 1565
QY 98 SerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeu 117
DB 1566 ACCTCGGAAACCTTAATGGTTTGTAGCCTCATTAAGCTTGACGAGTAGAAGATGTGG 1625
QY 118 GlyLeuGlu-----CysCys----- 122
DB 1626 GGTCAAGATAGCCATCAAAATATTGCTCATACGCGTGTGTACATGCGATTAGTGTAGG 1685
QY 123 ---LeuLeuTyrLeuSerLysThr 129
DB 1686 GAAGCGATGTATCTCCCTTTTACA 1709

RESULT 8
AAC38626/c
ID AAC38626 standard; DNA; 532 BP.
XX
AC AAC38626;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 21658.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX

PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.

PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.: 1,19
Score: 83.50
Percent Similarity: 41.48%
Best Local Similarity: 28.89%
Query Match: 12.03%
DB: 3
Length: 532
Matches: 39
Conservative: 17
Mismatch: 52
Indels: 27
Gaps: 7

US-09-974-546C-86 (1-135) x AAC38626 (1-532)

QY 10 TyrGluAspMetHisAsnIleHisIleLeuGlnIleArgLysLeuArgHisArgLeu 29
|||
445 TACAGCGACAGATGTGTTTTTGTGTCTCTCTTTATCCACTCAAGTAACCATCATCAC 386
DB 30 SerAsnPheProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCys 49

Db 385 TATAACTTACCAATCTAGTTGGTTGGCTTCCACACGGAACA-----TGT 338
QY 50 TyrLysValPheArgLysLysGluLysValLysAatqSerGlnLysAlaThrGluPhe--- 68
Db 337 AGCCATGGGTGCTCTTCCAGAAAGGTAGAGCGTCTCTTCATGCGCTCCCGAGTTGTA 278
QY 69 -----IleAspTyrSerIleGluGlnSerHisHisAla-----IleLeuThr--- 82
Db 277 GTTTTGTCTATACTTACCATCTCAGCTCGTCAACGCTGCAGATGCTGACCGGC 218
QY 83 -----ProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSer 98
Db 217 CAACCCGTGAACCCGACCATCATCTGACCTCCATTTCCCCAGTACCTCGTGTCTC 158
QY 99 SerLeu-----SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGln 115
Db 157 TCTACTCTCTGACTCCCTCACATCTCATCATGTTTCACTATG----- 116
QY 116 ThrLeuGlyLeuGluCysCys-----LeuLeuTyrLeuSer 127
Db 115 TCTCTGTCTACAGATGTTGTTCTGCGCTGATCTCTGCTTATCA 71

RESULT 9
ABZ80817/c
ID ABZ80817 standard; DNA; 146547 BP.
XX AC ABZ80817;
XX DT 13-JUN-2003 (first entry)
XX DE Human phospholipase C gamma 1 polymorphism G329ul.
XX KW human; single nucleotide polymorphism; SNP; phospholipase C gamma 1;
KW PLCG1; gene; vascular disease; plasminogen activator inhibitor type 2;
KW PAI-2; diagnosis; atherosclerosis; coronary artery disease; ischemia;
KW myocardial infarction; stroke; thromboembolism; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT variation replace(64001,c)
FT /*tag= a
XX WO2003007801-A2.
XX PN 30-JAN-2003.
XX PD 19-JUL-2002; 2002WO-US023041.
XX PF 20-JUL-2001; 2001US-0306941P.
XX PR 28-AUG-2001; 2001US-0315572P.
XX PR 05-OCT-2001; 2001US-0327488P.
XX PR 14-DEC-2001; 2001US-00017128.
XX PA (VITI-) VITIVITY INC.
XX PI McCarthy J;
XX WPI; 2003-278312/27.
XX
PT Diagnosing and prognosing vascular disease, e.g. coronary artery disease
PT and myocardial infarction, based on the detection of polymorphisms in the
PT phospholipase C gamma 1 and plasminogen activator inhibitor type 2 genes
PT and polypeptides.
XX
PS Claim 44; Page 204-245; 295pp; English.
XX
CC The invention relates to methods, agents and apparatus for diagnosing and
CC prognosing vascular diseases based on the detection of polymorphisms in
CC the phospholipase C gamma 1 (PLCG1) and plasminogen activator inhibitor
CC type 2 (PAI-2) genes and polypeptides. In a preferred claim of the
CC invention the polymorphism is a C to T nucleotide change at position

CC 64001 of the PLCG1 gene (reference sequence GI 11345540) causing an amino
CC acid change from Threonine to Isoleucine at position 813 of the protein.
CC This sequence represents the region of the PLCG1 gene with the T form of
CC the polymorphism. The methods, agents and apparatus is used for
CC diagnosing and prognosing vascular diseases e.g. atherosclerosis,
CC coronary artery disease (especially), myocardial infarction (especially),
CC ischemia, stroke, peripheral vascular diseases, venous thromboembolism
CC and pulmonary embolism
XX
SQ Sequence 146547 BP; 39552 A; 32625 C; 33339 G; 41031 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.69e+03 Length: 146547
Score: 83.00 Matches: 19
Percent Similarity: 65.71% Conservative: 4
Best Local Similarity: 54.23% Mismatches: 10
Query Match: 11.98% Indels: 2
DB: 8 Gaps: 1

US-09-974-546C-86 (1-135) x ABZ80817 (1-146547)
QY 16 IleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeu 35
Db 29020 ATTATCCCCCATTTTCAGATGAAAAAACTGAGGCACAGACTAAAGGACGAGATTCAAAC 28961
QY 36 ProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyr 50
Db 28960 CCAGGTAGTCTGGCCCCCAGATCAGTGTG-----TTTGTCTTC 28922

RESULT 10
AAAX13160/c
ID AAAX13160 standard; DNA; 24601 BP.
XX AC AAAX13160;
XX DT 19-MAR-1999 (first entry)
XX DE Enterococcus faecalis genome contig SEQ ID NO:223.
XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX OS Enterococcus faecalis.
XX PN WO9850555-A2.
XX PD 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US008985.
XX PR 06-MAY-1997; 97US-0044031P.
XX PR 16-MAY-1997; 97US-0046655P.
XX PR 14-NOV-1997; 97US-0066009P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Kunsch CA, Dillon PJ, Barash SC;
XX WPI; 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides -
PT used to develop products for the detection of Enterococcus and for use in
PT vaccines for prevention or attenuation of Enterococcus infection.
XX
PS Claim 1; Page 1125-1137; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAAX12938 to AAAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence of
CC Enterococcus faecalis in samples. They can also be used for diagnosing

CC Enterococcal infection in an animal and monitoring progression of
 CC disease, and for identifying agents which can be used to modulate the
 CC growth or pathogenicity of *Enterococcus faecalis*, or another related
 CC organism, in vivo or in vitro. In particular the polypeptides encoded by
 CC the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to
 CC prevent or attenuate an Enterococcal infection
 XX
 SQ Sequence 24601 BP; 8032 A; 4435 C; 5259 G; 6859 T; 0 U; 16 Other;

Alignment Scores:
 Pred. No.: 472 Length: 24601
 Score: 82.00 Matches: 42
 Percent Similarity: 47.10% Conservative: 23
 Best Local Similarity: 30.43% Mismatches: 51
 Query Match: 11.82% Indels: 22
 DB: 2 Gaps: 8

US-09-974-546C-86 (1-135) x AAX13160 (1-24601)

QY 7 AsnGlnLysTyrGluAsp-----MetHisAsnIlelle----- 17
 Db 2571 AACCAAAATCAAAATCAACTGAACTTTTGAAGACATTTTTTTCAGCAGCTTAATAAACAAC 2512
 QY 18 -----HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPhe 32
 Db 2511 GCTTAAATTAAGGCACATTCCTCTAAAGAAAGCCAGTAAATTAATAAGCCACCCGTT 2452
 QY 33 ProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysVal 52
 Db 2451 CCATCATTAATAATAATAATTCGACCACACCGCCCTCTCTTAATACTATACTATA 2392
 QY 53 PheArgLysGlyLysValLysArgSerGlnLysAlaThrGluPheLeuSerTyrSer 72
 Db 2391 AATCAAGAAAGAAAGAAATGAAGCTTCTCAAAA---AGAGGCTTCATT---TTTCT 2338
 QY 73 IleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
 Db 2337 ATT---CATTAAACCGCAGGAATATTGGCCCTTTAGAAATTTTCTGTATGATTCAGCA 2281
 QY 93 SerSerMetLysCysSerLeuSerSerGluAlaIleLeuThrHisLeuThrMetLysGly 110
 Db 2280 ACTTCTGCTGTCACA---ACTTGAATGCGTAATGTTTACTTCTCTTTATCT 2173

RESULT 11

AB98955/c
 ID ABS98955 standard; DNA; 24601 BP.

XX AC ABS98955;

XX DT 18-DEC-2002 (first entry)

XX DE Enterococcus faecalis contig sequence #223.

XX KW Computer readable medium; Enterococcus faecalis; microbe; growth;
 KW pathogenicity; vaccine; resistance; Enterococcal infection; commercial;
 KW therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;
 KW biotech technology; antibacterial; modulator of nucleic acid expression;
 KW contig; ds.
 XX OS Enterococcus faecalis.
 XX PN US2002120116-A1.
 XX PD 29-AUG-2002.
 XX PF 04-MAY-1998; 98US-00070927.
 XX PR 04-MAY-1998; 98US-00070927.

PA (KUNS/) KUNSCH C A.
 PA (DILL/) DILLON P J.
 XX (BARA/) BARASH S.
 PI Kunsch CA, Dillon PJ, Barash S;
 XX WPI; 2002-750065/81.
 XX
 PT Computer readable medium having recorded on it a *Enterococcus faecalis*
 PT nucleotide sequence useful for detecting diseases related to *Enterococcus*
 PT infections in animals.
 XX
 PS Claim 1; Page; 119pp; English.
 XX
 CC The present invention relates to a new computer readable medium with an
 CC Enterococcus faecalis nucleotide sequence. The invention is useful to
 CC diagnose the presence of *E. faecalis* in a sample or determining the
 CC presence of a specific microbe in a sample. The invention is also useful
 CC for modulating the growth or pathogenicity of *E. faecalis*, in a vaccine to
 CC confer resistance to Enterococcal infection, for commercial, therapeutic
 CC and industrial purposes, and for fermenting a particular sugar source or
 CC to produce a particular metabolite. The invention is useful for detecting
 CC diseases related to Enterococcus infections in animals, and for detecting
 CC *E. faecalis* using biotech technology. The present nucleic acid sequence
 CC represents an Enterococcus faecalis contig DNA sequence of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from USPTO
 CC at <http://seqdata.uspto.gov>
 XX
 SQ Sequence 24601 BP; 8032 A; 4435 C; 5259 G; 6859 T; 0 U; 16 Other;

Alignment Scores:

Pred. No.: 472 Length: 24601
 Score: 82.00 Matches: 42
 Percent Similarity: 47.10% Conservative: 23
 Best Local Similarity: 30.43% Mismatches: 51
 Query Match: 11.82% Indels: 22
 DB: 2 Gaps: 8

US-09-974-546C-86 (1-135) x ABS98955 (1-24601)

QY 7 AsnGlnLysTyrGluAsp-----MetHisAsnIlelle----- 17
 Db 2571 AACCAAAATCAAAATCAACTGAACTTTTGAAGACATTTTTTTCAGCAGCTTAATAAACAAC 2512
 QY 18 -----HisIleLeuGlnIleArgLysLeuArgHisArgLeuSerAsnPhe 32
 Db 2511 GCTTAAATTAAGGCACATTCCTCTAAAGAAAGCCAGTAAATTAATAAGCCACCCGTT 2452
 QY 33 ProArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysVal 52
 Db 2451 CCATCATTAATAATAATAATTCGACCACACCGCCCTCTCTTAATACTATACTATA 2392
 QY 53 PheArgLysGlyLysValLysArgSerGlnLysAlaThrGluPheLeuSerTyrSer 72
 Db 2391 AATCAAGAAAGAAAGAAATGAAGCTTCTCAAAA---AGAGGCTTCATT---TTTCT 2338
 QY 73 IleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
 Db 2337 ATT---CATTAAACCGCAGGAATATTGGCCCTTTAGAAATTTTCTGTATGATTCAGCA 2281
 QY 93 SerSerMetLysCysSerLeuSerSerGluAlaIleLeuPheThrLeuThr----- 110
 Db 2280 ACTTCTGCTGTCAGCT---TCTTTTAAACCCCTGATTTTTCGCTATCTCTTTTCG 2224
 QY 111 ---LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrLeuSer 127
 Db 2223 CCTTTCGCTGTCACA---ACTTGAATGCGTAATGTTTACTTCTCTTTATCT 2173

RESULT 12

ABD32622/c

ID ABD32622 standard; DNA; 133462 BP.

XX

AC	ABD32622;	Score: 81.50	Matches: 37
XX	18-NOV-2004 (first entry)	Percent Similarity: 42.86%	Conservative: 17
XX	Mouse cancer-associated genomic DNA MD13-011.	Best Local Similarity: 29.37%	Mismatches: 33
XX	Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;	Query Match: 11.74%	Indels: 39
XX	leukaemia; lymphoma; CAP.	DB: 13	Gaps: 8
XX	Mus musculus.	US-09-974-546C-86 (1-135) x ABD32622 (1-133462)	
KW	WO2004074330-A2.	QY 17 lleHislleLeuGlnlleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuPro 36	
XX	02-SEP-2004.	Db 9988 TCATACGCACTGCCTAAATGGTGATGTCAGATTCGGTTTTTACTAACTTCTCCAAA----- 9935	
XX	17-FEB-2004; 2004WO-US004730.	QY 37 GlylleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLys 56	
XX	14-FEB-2003; 2003US-00367094.	Db 9934 -----AATACTGTAAAGGCTCTCTTTCGAGGAAGTTA-----AAA 9899	
PR	15-MAR-2003; 2003US-00388838.	QY 57 GluLysValLysArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSer 76	
PR	15-APR-2003; 2003US-00417375.	Db 9898 AAAAAAAGGTCGTTCAAGTCAGTCAATGGCTTCTTACTTCTTAAATTCAGT 9839	
PR	13-JUN-2003; 2003US-00461862.	QY 77 HisHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLys 96	
PR	15-SEP-2003; 2003US-00663431.	Db 9838 AATTATGCTGTT-----AGCGGTTCGTC---CAA 9812	
PR	15-DEC-2003; 2003US-00737318.	QY 97 CysSerSerLeuSerSerGlu-----AlaIleLeuPheThrLeuThr 110	
XX	(SAGR-) SAGRES DISCOVERY INC.	Db 9811 TTTTCCTCTAAAGGAACAGTTTTTAAATATATAATATTCTCTTTTAAATATCTTT 9752	
XX	Morris DW, Morris DW, Malandro MS;	QY 111 LeuGlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeuTyrSerLysThrIle 130	
XX	WPI; 2004-652914/63.	Db 9751 TTAAGG-----AAGCTTTAT---TCAAAAAACAATA 9725	
XX	New isolated cancer-associated polynucleotides and polypeptides useful	QY 131 HisPro---GlnIleIle 135	
PT	for diagnosing, preventing or treating cancers, especially lymphoma and	Db 9724 CATCCCTCCCAAGTCATC 9707	
PT	leukemia, or in screening for agents that modulate cancer.	RESULT 13	
XX	disclosure; seqid 140; 310pp; English.	ABX44573	
XX	The invention relates to an isolated nucleic acid comprising at least 10	ID ABX44573 standard; cDNA; 381 BP.	
CC	contiguous nucleotides of any of the 233 polynucleotide sequences given	XX ABX44573;	
CC	in the specification, or its complement. The nucleic acids encode cancer-	AC	
CC	associated proteins. Also included are an expression vector comprising	XX	
CC	the isolated nucleic acid cited above, a host cell comprising the above	DT 21-FEB-2003 (first entry)	
CC	recombinant nucleic acid or expression vector, a microarray for detecting	DE Bovine EST associated with lactation/muscle/fat deposition #9738.	
CC	a cancer-associated (CA) nucleic acid comprising at least one probe	KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;	
CC	comprising at least 10 contiguous nucleotides of any of the above-	KW muscle deposition; fat deposition; genome mapping; gene identification;	
CC	mentioned nucleotide sequences, an isolated polypeptide (encoded within	XX gene analysis; cattle breeding.	
CC	an open reading frame of a CA sequence selected from any of the 95	OS Bos Taurus.	
CC	polynucleotide sequences as mentioned in the specification, or its	XX US2002137139-A1.	
CC	complement), an isolated antibody, (or its antigen binding fragment) that	PN	
CC	binds to the above polypeptide, a hybridoma that produces the above	PD 26-SEP-2002.	
CC	monoclonal antibody, a pharmaceutical composition comprising the above	XX	
CC	antibody and a pharmaceutical excipient, a kit for detecting cancer	PF 24-SEP-2001; 2001US-00960352.	
CC	cells(comprising the antibody cited above, methods for diagnosing cancer	PR 12-JAN-1999; 99US-0115707P.	
CC	or for detecting the presence or absence of cancer cells in an	PR 11-JAN-2000; 2000US-00480902.	
CC	individual, a method for inhibiting growth of cancer cells in an	XX (BYAT/) BYATT J C.	
CC	individual, a method for delivering a therapeutic agent to cancer cells	PA (MATH/) MATHIALAGAN N.	
CC	in an individual, an electronic library comprising the above	PA (TAON/) TAO N.	
CC	polynucleotide or polypeptide (or their fragments), methods of screening	PA (WARR/) WARREN W C.	
CC	for anticancer activity or for a bioactive agent capable of modulating	XX	
CC	the activity of a CA protein (CAP), methods for detecting cancer	PI Byatt JC, Mathialagan N, Tao N, Warren WC;	
CC	associated with expression of a polypeptide in a test cell sample, a	XX WPI; 2003-110599/10.	
CC	method for treating cancers and a method for inhibiting the expression of	XX	
CC	CA gene in a cell. The composition and methods are useful for detecting,	XX	
CC	diagnosing, preventing and treating cancers, especially lymphoma and	PT New nucleic acid associated with lactation, and muscle and fat	
CC	leukaemia. These may also be used in screening for agents that modulate	PT deposition, useful for genome mapping, gene identification and analysis,	
CC	cancer. The present sequence is a mouse CAP genomic sequence. Note: The	PT cattle breeding, or for genetically improving cattle.	
CC	sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences		
XX	Sequence 133462 BP; 39091 A; 24305 C; 25093 G; 41557 T; 0 U; 3416 Other;		
SQ			
Alignment Scores:			
Pred. No.:	6.32e-03	Length:	133462

XX

PS Claim 2; SEQ ID NO 9738; 245pp; English.

XX

CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived from
CC cattle, and the LMFD nucleic acid can specifically hybridize to a second
CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC appearing as ABX34836-ABX49947, or complements of them. Also included are
CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC acid linked to a promoter and a 3' non-translated sequence that
CC functions in the cell to cause termination of transcription and addition
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC (2) determining a level or pattern of a molecule in a bovine cell or
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC of the 15112 nucleic acid sequences or its complement or fragment) with a
CC complementary nucleic acid molecule obtained from the bovine cell or
CC tissue, where hybridisation between the marker nucleic acid and the
CC complementary nucleic acid permits the detection of the molecule; and (b)
CC detecting the level or pattern of the complementary nucleic acid; where
CC the detection of the complementary nucleic acid is predictive of the
CC level or pattern of the molecule. The LMFD nucleic acid is used for
CC determining a level or pattern of a molecule in a bovine cell or tissue.
CC It is useful for genome mapping, gene identification and analysis, cattle
CC breeding, preparation of constructs for use in cattle gene expression, or
CC for genetically improving cattle. The present sequence is one of the
CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC present sequence was not shown in the specification but was obtained in
CC electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139

XX

SQ Sequence 381 BP; 104 A; 84 C; 82 G; 111 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.51	Length:	381
Score:	81.00	Matches:	24
Percent Similarity:	42.55%	Conservative:	16
Best Local Similarity:	25.53%	Mismatches:	30
Query Match:	11.67%	Indels:	24
DB:	8	Gaps:	4

US-09-974-546C-86 (1-135) x ABX44573 (1-381)

QY 24 LysLeuArgHisArgLeuSerAsnPro-----ArgLeuProGly----- 37

DB 96 AAGATAAGACACAGAGATCTCAAACTGCCAACCTAGTCTGCTGCCAGTAGTGACGACGG 155

QY 38 -----lleLeuAlaProGluThrValLeuLeuProPheCys 49

DB 156 AAGGTTTGCCAAACCGTTGCGCCCGTACACCTGCCCGTTCACGATGCAACTGCTTCT 215

QY 50 TyrLysValPheArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPheIle 69

DB 216 GCAAAAGTGTACGATCGGAGGACTATGTC-----TTCAAT 251

QY 70 AspTyrSerIleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThr 89

DB 252 GAATTTTGTGTGCAACAGCTCAACATGTTAATGTAGATCCTTTAAATCGAAACTTACA 311

QY 90 MetLys-----GlySerMetLysCysSerSerLeuSer 101

DB 312 TTCAGTTGTCTTGGAGGAGTGTGTAATTTGTAACATTTAAAT 353

RESULT 14

ADN46845_11/c

Continuation (12 of 21) of ADN46845 from base 1100001 (Thermococcus kodakaraensis KOD1 9
WP Sequence split into 21 fragments LOCUS ADN46845 Accession Adn46845

WP	Fragment Name	Begin	End
WP	ADN46845_00	1	110000
WP	ADN46845_01	100001	210000
WP	ADN46845_02	200001	310000
WP	ADN46845_03	300001	410000
WP	ADN46845_04	400001	510000
WP	ADN46845_05	500001	610000

WP	ADN46845_06	600001	710000
WP	ADN46845_07	700001	810000
WP	ADN46845_08	800001	910000
WP	ADN46845_09	900001	1010000
WP	ADN46845_10	1000001	1110000
WP	ADN46845_11	1100001	1210000
WP	ADN46845_12	1200001	1310000
WP	ADN46845_13	1300001	1410000
WP	ADN46845_14	1400001	1510000
WP	ADN46845_15	1500001	1610000
WP	ADN46845_16	1600001	1710000
WP	ADN46845_17	1700001	1810000
WP	ADN46845_18	1800001	1910000
WP	ADN46845_19	1900001	2010000
WP	ADN46845_20	2000001	2089378

Alignment Scores:

Pred. No.:	5.52e+03	Length:	110000
Score:	81.00	Matches:	31
Percent Similarity:	44.36%	Conservative:	28
Best Local Similarity:	23.31%	Mismatches:	42
Query Match:	11.67%	Indels:	32
DB:	12	Gaps:	7

US-09-974-546C-86 (1-135) x ADN46845_11 (1-110000)

QY 13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSer----- 30

DB 68597 ATGCACAGCTT-----GCAATCTATCTGAAAAAGCTCGGCCACGCTTTTCAATAGTG 68544

QY 31 -----AsnPro 33

DB 68543 ACTAATGATCTTAAAAACGGGAAGAAAAAGAGCTTGAAGAGCTGGGGTAGGATTGGTT 68484

QY 34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53

DB 68483 AAGTCCCGGGTGTCTCATCGTCG---GTGCTCGGGATAAACATACCTAC----- 68436

QY 54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe---IleAspTyrSer 72

DB 68435 -----GGGCTGAAGTCGAACAGAGAACTAGGCGAGTTTCTCGTGGATTTCAT 68388

QY 73 IleGluGlnSerHisAlaIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92

DB 68387 GTGTCCAGCCGCCACACGCC---TTCACCTCCACTCTCGTGAAGCGGTTAAGCCGCG 68331

QY 93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111

DB 68330 AGAAGTCTCGAAAGGCAACACTCTCTCACACCCACAGCATATCTCTTCCACGAGTCT 68271

QY 112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeu 124

DB 68270 TCCCTATGGAAGGCCCTTGGGTGTGATCTTTTCCACTCTCTC 68232

RESULT 15

ADN47591_09

Continuation (10 of 21) of ADN47591 from base 900001 (Thermococcus kodakaraensis KOD1 9
WP Sequence split into 21 fragments LOCUS ADN47591 Accession Adn47591

WP	Fragment Name	Begin	End
WP	ADN47591_00	1	110000
WP	ADN47591_01	100001	210000
WP	ADN47591_02	200001	310000
WP	ADN47591_03	300001	410000
WP	ADN47591_04	400001	510000
WP	ADN47591_05	500001	610000
WP	ADN47591_06	600001	710000
WP	ADN47591_07	700001	810000
WP	ADN47591_08	800001	910000
WP	ADN47591_09	900001	1010000
WP	ADN47591_10	1000001	1110000
WP	ADN47591_11	1100001	1210000
WP	ADN47591_12	1200001	1310000
WP	ADN47591_13	1300001	1410000

WP ADN47591_14 1400001 1510000
WP ADN47591_15 1500001 1610000
WP ADN47591_16 1600001 1710000
WP ADN47591_17 1700001 1810000
WP ADN47591_18 1800001 1910000
WP ADN47591_19 1900001 2010000
WP ADN47591_20 2000001 2089378

Alignment Scores:

Pred. No.: 5.52e+03 Length: 110000
Score: 81.00 Matches: 31
Percent Similarity: 44.36% Conservative: 28
Best Local Similarity: 23.31% Mismatches: 42
Query Match: 11.67% Indels: 32
DB: 12 Gaps: 7

US-09-974-546C-86 (1-135) x ADN47591_09 (1-110000)

QY 13 MetHisAsnIleIleHisIleLeuGlnIleArgLysLeuArgHisArgLeuSer----- 30
Db 20781 ATGCACGAGCTT-----GCAATCTATCTGAAAAAGCTCGGCCACGACGCTTCAATAGTG 20834
QY 31 -----AsnPhePro 33
Db 20835 ACTAATGATCTAAACGCGGAAAAAGAGCTTGAAGAGCTGGGGTAGGATTGGTT 20894
QY 34 ArgLeuProGlyIleLeuAlaProGluThrValLeuLeuProPheCysTyrLysValPhe 53
Db 20895 AAGTCCCGGGTGTCATCAGTCCG---GTGCTCGGGATAACATACCTAC----- 20942
QY 54 ArgLysLysGluLysValLysArgSerGlnLysAlaThrGluPhe---IleAspTyrSer 72
Db 20943 -----GGGCTGAAGTCGACAGACAGAACTAGGCGAGTTTCTCGTGATTTTGAT 20990
QY 73 IleGluGlnSerHisHisIleIleLeuThrProLeuGlnThrHisLeuThrMetLysGly 92
Db 20991 GTGTCCACGCCACACGCC---TTCACTCCACTCTCGCTGAAGCGGTTAAGGCCGGG 21047
QY 93 SerSerMet---LysCysSerSerLeuSerSerGluAlaIleLeuPheThrLeuThrLeu 111
Db 21048 AGAAGCTCTGAAAAGGCAACACTCTCACACCCACAGCATATCTCTTCTCCACGAGTCT 21107
QY 112 GlnLeuThrGlnThrLeuGlyLeuGluCysCysLeuLeu 124
Db 21108 TCCCTATGGAAGGCCCTTGGGTTGACTTTTCCACTCCTC 21146

Search completed: September 9, 2005, 13:25:08
Job time : 388 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 13:12:05 ; Search time 104.5 Seconds
(without alignments)
2113.850 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 WRAFLRNQYEDMHHIHL.....TLGLSCCLLYLSKTHIPQII 135

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cg2_1/USPTO spoof/US09974546/runat.07092005.174503.21006/app_query.fasta_1.654
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=BITS -START=1 -END=1 -MATRIX=blosum62 -FRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09974546 @CGN 1.1.77 @runat.07092005.174503.21006 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cg2_6/prodata/1/ina/5A COMB.seq:
2: /cg2_6/prodata/1/ina/5B COMB.seq:
3: /cg2_6/prodata/1/ina/6A COMB.seq:
4: /cg2_6/prodata/1/ina/6B COMB.seq:
5: /cg2_6/prodata/1/ina/6C COMB.seq:
6: /cg2_6/prodata/1/ina/backfile1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	2087	3	US-09-097-199-83 Sequence 83, Appl
2	694	100.0	2505	3	US-09-097-199-85 Sequence 85, Appl
3	694	100.0	2506	4	US-09-949-016-3311 Sequence 3311, Ap
4	694	100.0	6507	4	US-09-949-016-15053 Sequence 15053, A
5	689	99.3	601	4	US-09-949-016-119365 Sequence 119365,
6	689	99.3	601	4	US-09-949-016-119366 Sequence 119366,
7	480	69.2	601	4	US-09-949-016-119364 Sequence 119364,
8	83.5	12.0	2472	4	US-09-248-796A-148 Sequence 148, App
9	80.5	11.6	1093	3	US-09-186-276B-53 Sequence 53, Appl
10	80.5	11.6	1093	3	US-08-842-445-53 Sequence 53, Appl
11	80.5	11.6	1093	3	US-09-186-188B-53 Sequence 53, Appl
12	80.5	11.6	1093	4	US-09-265-585C-53 Sequence 53, Appl

c 13	80	11.5	148783	4	US-09-949-016-15729	Sequence 15729, A
c 14	79.5	11.5	262	3	US-09-573-080A-99	Sequence 99, Appl
c 15	79.5	11.5	6196	3	US-08-675-566-5	Sequence 5, Appl
c 16	79.5	11.5	6243	3	US-08-675-566-14	Sequence 14, Appl
c 17	79.5	11.5	6503	3	US-08-675-566-6	Sequence 6, Appl
c 18	79.5	11.5	7379	3	US-08-675-566-13	Sequence 13, Appl
c 19	79.5	11.5	8618	3	US-08-675-566-21	Sequence 21, Appl
c 20	79.5	11.5	8792	3	US-08-675-566-25	Sequence 25, Appl
c 21	78	11.2	761	4	US-09-270-767-1826	Sequence 1826, Ap
c 22	78	11.2	761	4	US-09-270-767-17108	Sequence 17108, A
c 23	78	11.2	96878	4	US-09-949-016-12951	Sequence 12951, A
c 24	78	11.2	462589	4	US-09-949-016-12900	Sequence 12900, A
c 25	78	11.2	476044	4	US-09-949-016-12412	Sequence 12412, A
c 26	77	11.1	1461	1	US-08-587-670A-1	Sequence 1, Appl
c 27	77	11.1	1461	3	US-09-061-674-1	Sequence 1, Appl
c 28	77	11.1	22372	4	US-09-949-016-17459	Sequence 17459, A
c 29	76	11.0	367	3	US-09-328-111-446	Sequence 446, App
c 30	76	11.0	601	4	US-09-949-016-202994	Sequence 202994, A
c 31	76	11.0	42672	4	US-09-949-016-17253	Sequence 17253, A
c 32	76	11.0	42672	4	US-09-949-016-17254	Sequence 17254, A
c 33	76	11.0	66213	4	US-09-949-016-11803	Sequence 11803, A
c 34	76	11.0	66213	4	US-09-949-016-16739	Sequence 16739, A
c 35	75.5	10.9	19008	4	US-09-949-016-12923	Sequence 12923, A
c 36	75.5	10.9	160759	4	US-09-949-016-16514	Sequence 16514, A
c 37	75	10.8	64291	4	US-09-949-016-16278	Sequence 16278, A
c 38	75	10.8	117410	4	US-09-949-016-12262	Sequence 12262, A
c 39	74.5	10.7	3915	4	US-09-023-655-1104	Sequence 1104, Ap
c 40	74.5	10.7	5522	4	US-09-949-016-1259	Sequence 1259, Ap
c 41	74.5	10.7	5523	4	US-09-949-016-1014	Sequence 1014, Ap
c 42	74.5	10.7	27120	4	US-09-949-016-16210	Sequence 16210, A
c 43	74.5	10.7	35707	4	US-09-949-016-17120	Sequence 17120, A
c 44	74	10.7	601	4	US-09-949-016-157153	Sequence 157153, A
c 45	74	10.7	3143	4	US-09-949-016-856	Sequence 856, App

ALIGNMENTS

RESULT 1
US-09-097-199-83
; Sequence 83, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltre, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:

```
/
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 83:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2087 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 99..503
/
US-09-097-199-83
/
Alignment Scores:
Pred. No.: 1,65e-85 Length: 2087
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-974-546C-86 (1-135) x US-09-097-199-83 (1-2087)
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCTTCTTAAGGAACAGAAATATAGGATATGCACAATATTATTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAAATTGAGGCACAGATTAACTTCCCAAGGCTACCGCATTCCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
Db 219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAGAAAAAGTAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 279 AGAAGTCAAAAGGCCAACAGAGTTCAATGATTATTCATAGAACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
Db 339 CTCACACCCCTTCGACACACATGACCATGAAGGTTCTCAATGAAATGTTCCCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 399 TCTTCAGAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCCTAGTCTGGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 459 TGCTGTCTTCTCTACTATTCCAAAACATATACATCCACAGATCAT 503

RESULT 2
US-09-097-199-85
/ Sequence 85, Application US/09097199
/ Patent No. 6218529
/ GENERAL INFORMATION:
/ APPLICANT: An, Gang
/ APPLICANT: O'Hara, S. Mark
/ APPLICANT: Ralph, David
/ APPLICANT: Veltri, Robert
/ TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
/ PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
/ NUMBER OF SEQUENCES: 87
/ CORRESPONDENCE ADDRESS:
/ ADDRESSES: Arnold, White & Durkee
/ STREET: P.O. Box 4433
/ CITY: Houston
/ STATE: Texas
/ COUNTRY: USA
/ ZIP: 77210
/
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```
/
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/097,199
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/692,787
/ FILING DATE: 31-JUL-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Nakashima, Richard A.
/ REGISTRATION NUMBER: P-42,023
/ REFERENCE/DOCKET NUMBER: UROC:018
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 85:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2505 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 99..503
/
US-09-097-199-85
/
Alignment Scores:
Pred. No.: 2,21e-85 Length: 2505
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-09-974-546C-86 (1-135) x US-09-097-199-85 (1-2505)
QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCTTCTTAAGGAACAGAAATATGAGATATGCACAATATTATTCACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAAATTGAGGCACAGATTAACTTCCCAAGGCTACCGCATTCCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysLysGluLysValLys 60
Db 219 CCAGAAACTGTGCTCTTACCATTCTGCTACAAGGTATTTCGAAAAAAGAAAAAGTAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 279 AGAAGTCAAAAGGCCAACAGAGTTCAATGATTATTCATAGAACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerLeu 100
Db 339 CTCACACCCCTTCGACACACATGACCATGAAGGTTCTCAATGAAATGTTCCCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 399 TCTTCAGAGCCATATTATTCACATTGACTTTGCAGTTAACTCAGACCCCTAGTCTGGAA 458
QY 121 CysCysLeuLeuTyrLeuSerLysThrIleHisProGlnIleIle 135
Db 459 TGCTGTCTTCTCTACTATTCCAAAACATATACATCCACAGATCAT 503

RESULT 3
US-09-949-016-3311
/ Sequence 3311, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 9, 2005, 14:46:00 ; Search time 408.5 Seconds
(without alignments)
2170.324 Million cell updates/sec

Title: US-09-974-546C-86

Perfect score: 694

Sequence: 1 MRAFLRNQYEDMHNIIHL.....TLGLECCLLYLSKTHPQII 135

Scoring table:

BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7351250 seqs, 3283620254 residues

Total number of hits satisfying chosen parameters: 14702500

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09974546/runat_07092005_174504_21077/app_query.fasta_1.654
-DB=Published Applications NA -OEMT=fastap -SUPFI=p2n.inpb -MINMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-MAXLEN=2000000000 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO MMAP -US09974546 -CGN 1.1 456 @runat_07092005_174504_21077
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FCGAPOP=6 -FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/2/pubpna/US10J_PUBCOMB.seq:
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	694	100.0	2087	10	US-09-974-546-83	Sequence 83, Appl
2	694	100.0	2505	10	US-09-974-546-85	Sequence 85, Appl
3	90	13.0	1113	17	US-10-369-493-44085	Sequence 44085, A
4	90	13.0	1146	17	US-10-369-493-25021	Sequence 25021, A
5	87	12.5	1258	18	US-10-425-114-318	Sequence 318, App
6	87	12.5	1624	20	US-10-425-115-49918	Sequence 49918, A
7	83	12.0	146547	15	US-10-017-128-1	Sequence 1, Appl
8	82	11.8	24601	9	US-09-070-927A-223	Sequence 223, App
9	81.5	11.7	133462	19	US-10-367-094-114	Sequence 114, App
10	81	11.7	381	9	US-09-960-352-9738	Sequence 9738, Ap
11	81	11.7	819	19	US-10-437-963-23499	Sequence 23499, A
12	81	11.7	1277	20	US-10-425-115-137953	Sequence 137953, A
13	81	11.7	4437	20	US-10-425-115-178334	Sequence 178334, A
14	80.5	11.6	455	20	US-10-425-115-63753	Sequence 63753, A
15	80.5	11.6	600	22	US-10-972-079-4637	Sequence 4637, Ap
16	80.5	11.6	600	22	US-10-972-079-4638	Sequence 4638, Ap
17	80.5	11.6	1026	18	US-10-425-114-13022	Sequence 13022, A
18	80.5	11.6	1093	9	US-09-186-2768-53	Sequence 53, Appl
19	80.5	11.6	1093	9	US-09-186-1888-53	Sequence 53, Appl
20	80.5	11.6	1093	14	US-10-253-007-53	Sequence 53, Appl
21	80.5	11.6	2343	20	US-10-425-115-127057	Sequence 127057, A
22	80	11.5	681	19	US-10-437-963-98199	Sequence 98199, A
23	80	11.5	1355	19	US-10-437-963-37754	Sequence 37754, A
24	80	11.5	1693	20	US-10-723-860-7296	Sequence 7296, Ap
25	79.5	11.5	262	10	US-09-854-867-99	Sequence 99, Appl
26	79.5	11.5	262	21	US-10-786-970A-99	Sequence 99, Appl
27	79.5	11.5	567	18	US-10-424-599-74400	Sequence 74400, A
28	79.5	11.5	585	20	US-10-357-930-54224	Sequence 54224, A
29	79	11.4	590	13	US-10-027-632-269701	Sequence 269701, A
30	79	11.4	590	13	US-10-027-632-269701	Sequence 269701, A
31	79	11.4	831	13	US-10-027-632-133030	Sequence 133030, A
32	79	11.4	831	17	US-10-027-632-133030	Sequence 133030, A
33	79	11.4	4068	19	US-10-437-963-35166	Sequence 35166, A
34	79	11.4	4107	20	US-10-425-115-139445	Sequence 139445, A
35	79	11.4	4335	20	US-10-425-115-139432	Sequence 139432, A
36	78.5	11.3	386	9	US-09-960-352-10443	Sequence 10443, A
37	78.5	11.3	600	22	US-10-972-079-88625	Sequence 88625, A
38	78.5	11.3	1787	13	US-10-027-632-84669	Sequence 84669, A
39	78.5	11.3	1787	13	US-10-027-632-84670	Sequence 84670, A
40	78.5	11.3	1787	17	US-10-027-632-84669	Sequence 84669, A
41	78.5	11.3	1787	17	US-10-027-632-84670	Sequence 84670, A
42	78	11.2	650	13	US-10-027-632-223621	Sequence 223621, A
43	78	11.2	650	17	US-10-027-632-223621	Sequence 223621, A
44	78	11.2	1207	18	US-10-424-599-5755	Sequence 5755, Ap
45	78	11.2	9869	20	US-10-425-115-139430	Sequence 139430, A

ALIGNMENTS

RESULT 1

US-09-974-546-83
; Sequence 83, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Veltri, Robert

TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,

PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE

NUMBER OF SEQUENCES: 87

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/974,546
FILING DATE: 10-Oct-2001
CLASSIFICATION: Unknown
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/097,199
FILING DATE: 1998-06-12
ATTORNEY/AGENT INFORMATION:
NAME: Nakashima, Richard A.
REGISTRATION NUMBER: P-42,023
REFERENCE/DOCKET NUMBER: UROC:018
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 2087 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 99..503
SEQUENCE DESCRIPTION: SEQ ID NO: 83:
US-09-974-546-83

Alignment Scores:
Pred. No.: 5,82e-85 Length: 2087
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-974-546C-86 (1-135) x US-09-974-546-83 (1-2087)

QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCCTCTTAAGGAACAGAAATATGAGGATATGCACAATATATTACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAAATTGAGGCACAGATTAAGTAATCTCCCAAGGCTACCAGGCATTTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60
Db 219 CCAGAAACTGTGCTCTTACCATTTGCTACAAGGTATTTGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80
Db 279 AGAAGTCAAAAGGCAACAGATTCATTGATTAATCCATAGACAGTCACACCATGCAATT 338
QY 81 LeuThrProLeuGlnThrHisLeuThrMetLysGlySerSerMetLysCysSerSerLeu 100
Db 339 CTCACACCCCTTCAGACACACTTGACCATGAAAGGTTCTCAATGAAATGTTCTCATTA 398
QY 101 SerSerGluAlaIleLeuPheThrLeuThrLeuGlnLeuThrGlnThrLeuGlyLeuGlu 120
Db 399 TCTTCAGAGCCATATATTACATTGACTTTGCAAGTTAACTCAGACCCCTAGGCTCGGAA 458
QY 121 CysCysLeuLeuTyrLysSerLysThrIleHisProGlnIleIle 135
Db 459 TGCTGTCTCTCTACTATTACAAACTATACATCCACAGATCATATA 503

RESULT 2

US-09-974-546-85
Sequence 85, Application US/09974546
Publication No. US20030050470A1
GENERAL INFORMATION:
APPLICANT: An, Gang
O'Hara, S. Mark
Ralph, David
Veltri, Robert
TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS
PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/974,546
FILING DATE: 10-Oct-2001
CLASSIFICATION: Unknown
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/097,199
FILING DATE: 1998-06-12
ATTORNEY/AGENT INFORMATION:
NAME: Nakashima, Richard A.
REGISTRATION NUMBER: P-42,023
REFERENCE/DOCKET NUMBER: UROC:018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 2505 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 99..503
SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-09-974-546-85

Alignment Scores:
Pred. No.: 7,74e-85 Length: 2505
Score: 694.00 Matches: 135
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-974-546C-86 (1-135) x US-09-974-546-85 (1-2505)

QY 1 MetArgAlaPheLeuArgAsnGlnLysTyrGluAspMetHisAsnIleIleHisIleLeu 20
Db 99 ATGAGGGCCCTCTTAAGGAACAGAAATATGAGGATATGCACAATATATTACATTTTA 158
QY 21 GlnIleArgLysLeuArgHisArgLeuSerAsnPheProArgLeuProGlyIleLeuAla 40
Db 159 CAGATCAGAAAATTGAGGCACAGATTAAGTAATCTCCCAAGGCTACCAGGCATTTCTAGCT 218
QY 41 ProGluThrValLeuLeuProPheCysTyrLysValPheArgLysGluLysValLys 60
Db 219 CCAGAAACTGTGCTCTTACCATTTGCTACAAGGTATTTGAAAAAAGAAAAAGTAAAA 278
QY 61 ArgSerGlnLysAlaThrGluPheIleAspTyrSerIleGluGlnSerHisAlaIle 80

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: September 9, 2005, 13:09:35 ; Search time 1867.5 Seconds
(without alignments)
2751.634 Million cell updates/sec

Title: US-09-974-546C-86
Perfect score: 694
Sequence: 1 MRAFLRNQYEDWHNIHL.....TLGLECLLYLSKTHPQII 135

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-O=/cgn2.1/USPTO.spool/US09974546/runat_07092005.174502.20989/app_query.fasta_1.654
-DB=EST -OPMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09974546 @CGN 1.1.3556 @runat_07092005.174502.20989 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gse1:*
9: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	97	14.0	690	5	BU242169
2	95.5	13.8	588	5	BW569427
3	94	13.5	606	6	CB286029
4	90.5	13.0	459	5	BW557636
5	90.5	13.0	543	5	BW582739
6	90	13.0	570	5	BW591794
7	89.5	12.9	1151	5	BU413225
8	89	12.8	811	5	BU436282
9	88.5	12.8	743	7	CV511381

10	88.5	12.8	789	8	BH559290
11	88	12.7	776	9	AG377362
C 12	87.5	12.6	803	5	BX756993
C 13	87	12.5	788	9	CR262699
C 14	86.5	12.5	741	3	AY439721
C 15	86	12.4	529	5	BQ592283
C 16	86	12.4	769	8	BZ521554
C 17	86	12.4	992	9	CL254017
C 18	85.5	12.3	788	8	CC091268
C 19	85.5	12.3	877	6	CA472302
C 20	85	12.2	579	9	CE464906
C 21	85	12.2	707	9	CC743728
C 22	85	12.2	906	9	CG300818
C 23	85	12.2	914	9	CG300807
C 24	84.5	12.2	285	4	BM111157
C 25	84.5	12.2	562	8	AZ954412
C 26	84.5	12.2	652	2	BF644675
C 27	84.5	12.2	688	5	BP027905
C 28	84.5	12.2	741	4	BI309709
C 29	84.5	12.2	770	4	BM780176
C 30	84	12.1	493	5	BX836326
C 31	84	12.1	624	8	BZ745987
C 32	84	12.1	728	6	CD821188
C 33	84	12.1	896	9	CG008627
C 34	84	12.1	916	7	CK151729
C 35	84	12.1	1399	2	BE739923
C 36	84	12.1	1399	2	AW730867
C 37	83.5	12.0	721	8	B69065
C 38	83.5	12.0	732	5	BU264381
C 39	83.5	12.0	781	7	CNI135272
C 40	83.5	12.0	952	1	AL665968
C 41	83.5	12.0	1335	4	BM467342
C 42	83	12.0	395	6	CB069753
C 43	83	12.0	572	2	BE015616
C 44	83	12.0	612	2	BE015598
C 45	83	12.0	634	9	AG110445

ALIGNMENTS

RESULT 1
BU242169/c
LOCUS BU242169 603779592F1 CSEQCHN34 Gallus gallus linear EST 26-NOV-2002
DEFINITION 603779592F1 CSEQCHN34 Gallus gallus clone Chest725h23 5', mRNA
sequence.
ACCESSION BU242169
VERSION BU242169.1 GI:25488447
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE 1 (bases 1 to 690)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445332
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. 690
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"

BH559290 BOH080TR
AG377362 Mus muscu
BX756993 BX756993
CR262699 Reverse s
AY439721 Armigeres
BQ592283 BQ12698-0
BZ521554 BOKA811F
CL254017 ZMMBB060
CC091268 CSU-K33r.
CA472302 AGENCOURT
CE464906 tigr-gss-
CC743728 ZMMBB011
CG300818 OG0A825TV
CG300807 OG0A825TH
BM111157 EST558693
AZ954412 2M022005
BF644675 NF018F01E
BP027905 BP027905
BI309709 EST51119
BM780176 EST590764
BX836326 BX836326
BZ745987 OGFBK05TM
CD821188 BN25.041B
CG008627 ZUAC085TH
CK151729 FGAS03441
BE739923 601556175
AW730867 GA_Ea002
B69065 CIT-HSP-205
BU264381 603505585
CNI135272 OX1_31_E0
AL665968 AL665968
BM467342 AGENCOURT
CB069753 1a28d02.X
BE015616 EST842.X
BE015598 EST824.Ma
AG110445 Pan trogl

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:23:48 ; Search time 58 Seconds
(without alignments)
900.218 Million cell updates/sec

Title: US-09-974-546C-86

Perfect score: 694

Sequence: 1 MRAPLRNQYEDWHNIHL.....TLGLECLLYLSTKTHPQII 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	694	100.0	135	3 AAY59296	Aay59296 Prostate
2	694	100.0	135	3 AAY59295	Aay59295 Prostate
3	694	100.0	135	4 AAU02174	Aau02174 Biomarker
4	694	100.0	135	4 AAU02175	Aau02175 Biomarker
5	103	14.8	21	3 AAY59294	Aay59294 Cancer bi
6	103	14.8	21	4 AAU02173	Aau02173 Biomarker
7	90	13.0	371	8 ADS41968	Ads41968 Bacterial
8	90	13.0	381	8 ADN18681	Adn18681 Bacterial
9	82.5	11.9	645	5 ABB90843	Abb90843 Herbicida
10	81	11.7	384	8 ADN48099	Adn48099 Thermococ
11	77.5	11.2	1061	7 ADF50279	Adf50279 Human PRM
12	77.5	11.2	1061	8 ADO36741	Ado36741 Human PR-
13	77.5	11.2	1144	4 ABB58124	Abb58124 Drosophil
14	77	11.1	1026	8 ADS08184	Ads08184 Staphyloc
15	76.5	11.0	322	6 ABM69043	Abm69043 Photorhab
16	75	10.8	633	6 ABU42279	Abu42279 Protein e
17	73	10.5	195	4 ABG18410	Abg18410 Novel hum
18	72	10.4	523	4 ABA46398	Ab46398 H. pylori
19	72	10.4	636	8 ADM44292	Adm44292 Soybean E
20	71.5	10.3	450	7 ADF41696	Adf41696 Bacillus
21	71.5	10.3	634	8 ADM94288	Adm94288 Corn Ethy
22	71.5	10.3	676	4 ABG15498	Abg15498 Novel hum
23	71	10.2	127	6 AAU42328	Aau42328 Propionib
24	71	10.2	127	6 ABM38847	Abm38847 Propionib
25	71	10.2	424	4 AAM39676	Aam39676 Human pol

26	71	10.2	424	4 AAG81358	Aag81358 Human AFP
27	71	10.2	424	8 ADH09599	Adh09599 Human hos
28	71	10.2	424	8 ADH09598	Adh09598 Human hos
29	71	10.2	424	8 ADH09600	Adh09600 Human hos
30	71	10.2	429	4 AAM41462	Aam41462 Human pol
31	71	10.2	519	4 AAB92584	Aab92584 Human pro
32	71	10.2	1922	8 ADL97802	Adl97802 Human dcr
33	71	10.2	1924	4 AAE09768	Aae09768 Human dic
34	71	10.2	1924	5 AAO15987	Aao15987 Human hel
35	71	10.2	1924	6 ABO07204	Abo07204 Human p53
36	71	10.2	1924	7 ABU64479	Abu64479 Stem loop
37	71	10.2	1924	7 ADN60159	Adn60159 Human hel
38	70.5	10.2	438	7 ABO64633	Abo64633 Klebsiell
39	70.5	10.2	634	8 ADM94300	Adm94300 Corn Ethy
40	70	10.1	206	7 ADJ69892	Adj69892 Human hea
41	70	10.1	982	8 ADS29846	Ads29846 Bacterial
42	69.5	10.0	1294	2 AAW30601	Aaw30601 Human typ
43	69.5	10.0	1353	8 ADQ89904	Adq89904 Antagonis
44	69.5	10.0	1398	3 AAB18292	Aab18292 Plasmodiu
45	69.5	10.0	3105	5 ABB80604	Abb80604 Human sbg

ALIGNMENTS

RESULT 1

AAY59296

ID AAY59296 standard; peptide; 135 AA.

XX AC AAY59296;

XX DT 19-APR-2000 (first entry)

XX DE Prostate disease marker UC Band #28 amino acid sequence.

XX KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;

XX KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.

XX OS Homo sapiens.

XX PN WO9964631-A1.

XX PD 16-DEC-1999.

XX PF 11-JUN-1999; 99WO-US013151.

XX PR 12-JUN-1998; 98US-00097199.

XX PA (UROC-) UROCOR INC.

XX PI An G, O'hara SM, Ralph D, Veltri RW;

XX DR WPI; 2000-116557/10.

XX DR N-PSDB; AA287584.

XX PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,

XX PT breast and bladder cancer.

XX PS Example 5; Page 184-186; 191pp; English.

XX CC The invention provides nucleic acid markers of prostate, breast and

XX CC bladder cancer. The markers are indicators of malignant transformation of

XX CC prostate, breast and bladder tissues and are diagnostic of the potential

XX CC for metastatic spread of malignant prostate tumours. The nucleic acid can

XX CC also be used as targets for therapeutic intervention in prostate cancer.

XX CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The

XX CC markers may be used to design specific probes and primers, for the rapid

XX CC analysis of prostate, bladder or breast biopsy samples. The probes and

XX CC primers may also be used for in situ hybridization or in situ PCR

XX CC detection and diagnosis. They may also be used to identify and isolate

XX CC full length gene sequences form various DNA libraries. Antibodies against

XX CC the polypeptide products of the markers can be used to treat prostate

XX CC cancer, bladder cancer or breast cancer. The encoded proteins may be used

CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
XX Sequence 135 AA;
SQ
Query Match 100.0%; Score 694; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9e-75;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRAFLRNQKYEDMHNIIHILOIRKLHRLSNFPRLPGLAPETVLLPFCYKVRKKEKVK 60
DB 1 MRAFLRNQKYEDMHNIIHILOIRKLHRLSNFPRLPGLAPETVLLPFCYKVRKKEKVK 60
QY 61 RSQKATEFDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
DB 61 RSQKATEFDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
QY 121 CCLLYLSKTIHPQII 135
DB 121 CCLLYLSKTIHPQII 135
RESULT 2
AAU02174
ID AAY59295 standard; peptide; 135 AA.
XX
AC AAY59295;
XX
DT 19-APR-2000 (first entry)
XX
DE Prostate disease marker UC Band #28 amino acid sequence.
XX
KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX
OS Homo sapiens.
XX
PN WO9664631-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US013151.
XX
PR 12-JUN-1998; 98US-00097199.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri RW;
XX
DR WPI; 2000-116557/10.
DR N-PSDB; AA287583.
XX
PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
PT breast and bladder cancer.
XX
PS Example 5; Page 182-183; 191pp; English.
XX
CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers, for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences form various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in

CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX
XX Sequence 135 AA;
SQ
Query Match 100.0%; Score 694; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9e-75;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRAFLRNQKYEDMHNIIHILOIRKLHRLSNFPRLPGLAPETVLLPFCYKVRKKEKVK 60
DB 1 MRAFLRNQKYEDMHNIIHILOIRKLHRLSNFPRLPGLAPETVLLPFCYKVRKKEKVK 60
QY 61 RSQKATEFDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
DB 61 RSQKATEFDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
QY 121 CCLLYLSKTIHPQII 135
DB 121 CCLLYLSKTIHPQII 135
RESULT 3
AAU02174
ID AAU02174 standard; protein; 135 AA.
XX
AC AAU02174;
XX
DT 29-AUG-2001 (first entry)
XX
DE Biomarker protein encoded by UC band 28 #2, used in diagnosis of cancer.
XX
KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human.
XX
OS Homo sapiens.
XX
PN US6218529-B1.
XX
PD 17-APR-2001.
XX
PF 12-JUN-1998; 98US-00097199.
XX
PR 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.
XX
PA (UROC-) UROCOR INC.
XX
PI An G, O'hara SM, Ralph D, Veltri R;
XX
DR WPI; 2001-289849/30.
DR N-PSDB; AAS04000.
XX
PT New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX
PS Disclosure; Col 121; 78pp; English.
XX
CC The sequence represents the amino acid sequence of biomarker protein
CC encoded by UC band 28 #2. Proteins encoded by the nucleic acid markers
CC can be used to produce antibodies for the detection of prostate, breast
CC or bladder cancer. Biomarker nucleic acid sequences can be used as
CC hybridisation probes and primers that specifically hybridise to prostate
CC cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast
CC cancer markers. The nucleic acids can be used as targets for therapeutic
CC intervention in these diseases, in the identification and isolation of
CC full-length gene sequences, including regulatory elements for gene
CC expression, from genomic human DNA libraries, as hybridisation probes for
CC screening genomic human DNA libraries. The kits comprising the nucleic
CC acid sequences are useful for detecting bladder, breast or prostate

CC cancer cells in a biological sample
XX SQ Sequence 135 AA;

Query Match 100.0%; Score 694; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9e-75;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNPPRLPGILAPETVLLPFCYKVRKKEKVK 60
DB 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNPPRLPGILAPETVLLPFCYKVRKKEKVK 60
QY 61 RSQKATEFIDYSEQSHHAILTPLQTLTKGSSMKCSSLSSEAILFTTLQTLQTLGLE 120
DB 61 RSQKATEFIDYSEQSHHAILTPLQTLTKGSSMKCSSLSSEAILFTTLQTLQTLGLE 120
QY 121 CCLLYLSKTIHPQII 135
DB 121 CCLLYLSKTIHPQII 135

RESULT 4
AAU02175
ID AAU02175 standard; protein; 135 AA.
XX AC AAU02175;
XX DT 29-AUG-2001 (first entry)
XX DE Biomarker protein encoded by UC band 28 #3, used in diagnosis of cancer.
XX KW Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
XX KW benign prostatic hyperplasia; BPH; therapeutic; human.
XX OS Homo sapiens.
XX PN US6218529-B1.
XX PD 17-APR-2001.
XX PF 12-JUN-1998; 98US-00097199.
XX PR 31-JUL-1995; 95US-0001655P.
XX PR 11-JAN-1996; 96US-0013611P.
XX PR 31-JUL-1996; 96US-00692787.
XX PA (UROC-) UROCOR INC.
XX PI An G, O'hara SM, Ralph D, Veltri R;
XX DR N-PSDB; AAS04001.
XX PT New nucleic acids as biomarkers and targets useful for detecting,
XX PT diagnosing, prognosing, and in developing treatments for prostate, breast
XX PT and bladder cancer.
XX PS Disclosure; Col 125; 78pp; English.

CC The sequence represents the amino acid sequence of biomarker protein
CC encoded by UC band 28 #3. Proteins encoded by the nucleic acid markers
CC can be used to produce antibodies for the detection of prostate, breast
CC or bladder cancer. Biomarker nucleic acid sequences can be used as
CC hybridisation probes and primers that specifically hybridise to prostate
CC cancer, benign prostatic hyperplasia (BPH), bladder cancer or breast
CC cancer markers. The nucleic acids can be used as targets for therapeutic
CC intervention in these diseases, in the identification and isolation of
CC full-length gene sequences, including regulatory elements for gene
CC expression, from genomic human DNA libraries, as hybridisation probes for
CC screening genomic human DNA libraries. The kits comprising the nucleic
CC acid sequences are useful for detecting bladder, breast or prostate
CC cancer cells in a biological sample
XX

CC cancer cells in a biological sample
XX SQ Sequence 135 AA;

Query Match 100.0%; Score 694; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 6.9e-75;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNPPRLPGILAPETVLLPFCYKVRKKEKVK 60
DB 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNPPRLPGILAPETVLLPFCYKVRKKEKVK 60
QY 61 RSQKATEFIDYSEQSHHAILTPLQTLTKGSSMKCSSLSSEAILFTTLQTLQTLGLE 120
DB 61 RSQKATEFIDYSEQSHHAILTPLQTLTKGSSMKCSSLSSEAILFTTLQTLQTLGLE 120
QY 121 CCLLYLSKTIHPQII 135
DB 121 CCLLYLSKTIHPQII 135

RESULT 5
AA59294
ID AA59294 standard; peptide; 21 AA.
XX AC AA59294;
XX DT 19-APR-2000 (first entry)
XX DE Cancer biomarker UC28 antigenic peptide.
XX KW Nucleic acid marker; biomarker; tumour; prostate cancer; bladder cancer;
XX KW benign prostatic hyperplasia; BPH; breast cancer; human; immunodetection.
XX OS Homo sapiens.
XX PN WO9964631-A1.
XX PD 16-DEC-1999.
XX PF 11-JUN-1999; 99WO-US013151.
XX PR 12-JUN-1998; 98US-00097199.
XX PA (UROC-) UROCOR INC.
XX PI An G, O'hara SM, Ralph D, Veltri RW;
XX DR WPI; 2000-116557/10.
XX PT Novel RNA biomarkers for diagnosis, prognosis and management of prostate,
XX PT breast and bladder cancer.
XX PS Example 5; Page 114; 191pp; English.

CC The invention provides nucleic acid markers of prostate, breast and
CC bladder cancer. The markers are indicators of malignant transformation of
CC prostate, breast and bladder tissues and are diagnostic of the potential
CC for metastatic spread of malignant prostate tumours. The nucleic acid can
CC also be used as targets for therapeutic intervention in prostate cancer,
CC benign prostatic hyperplasia (BPH), bladder cancer or breast cancer. The
CC markers may be used to design specific probes and primers for the rapid
CC analysis of prostate, bladder or breast biopsy samples. The probes and
CC primers may also be used for in situ hybridization or in situ PCR
CC detection and diagnosis. They may also be used to identify and isolate
CC full length gene sequences of various DNA libraries. Antibodies against
CC the polypeptide products of the markers can be used to treat prostate
CC cancer, bladder cancer or breast cancer. The encoded proteins may be used
CC to detect antibodies. The proteins and antibodies can be used in
CC immunodetection methods for detecting or quantifying the cancers, and for
CC clinical diagnosis of these cancers. The antibodies may also be used for
CC radioimaging to quantify and localize the encoded proteins
XX SQ Sequence 21 AA;

Query Match 14.8%; Score 103; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 RKKEKVRKSKATEFIDYSIE 74
DB 1 RKKEKVRKSKATEFIDYSIE 21

RESULT 6
AAU02173
ID AAU02173 standard; peptide; 21 AA.

XX AAU02173;
AC 29-AUG-2001 (first entry)
XX Biomarker UC band 28, antigenic peptide used in diagnosis of cancer.
XX Prostate; breast; bladder; cancer; biomarker; probe; diagnostic;
KW benign prostatic hyperplasia; BPH; therapeutic; human; antigenic.
XX Homo sapiens.
XX US6218529-B1.
XX 17-APR-2001.
XX 12-JUN-1998; 98US-00097199.
XX 31-JUL-1995; 95US-0001655P.
PR 11-JAN-1996; 96US-0013611P.
PR 31-JUL-1996; 96US-00692787.
XX (UROC-) UROCOR INC.

XX An G, O'hara SM, Ralph D, Veltri R;
XX WPI; 2001-289849/30.
XX New nucleic acids as biomarkers and targets useful for detecting,
PT diagnosing, prognosing, and in developing treatments for prostate, breast
PT and bladder cancer.
XX Example 5; Col 73; 78pp; English.

XX The sequence represents the amino acid sequence of biomarker, UC band 28,
CC antigenic peptide used to produce antibodies for the detection of
CC prostate, breast or bladder cancer. Biomarker nucleic acid sequences can
CC be used as hybridisation probes and primers that specifically hybridise
CC to prostate cancer, benign prostatic hyperplasia (BPH), bladder cancer or
CC breast cancer markers. The nucleic acids can be used as targets for
CC therapeutic intervention in these diseases, in the identification and
CC isolation of full-length gene sequences, including regulatory elements
CC for gene expression, from genomic human DNA libraries, as hybridisation
CC probes for screening genomic human DNA libraries. The kits comprising the
CC nucleic acid sequences are useful for detecting bladder, breast or
CC prostate cancer cells in a biological sample
XX Sequence 21 AA;

Query Match 14.8%; Score 103; DB 4; Length 21;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 RKKEKVRKSKATEFIDYSIE 74
DB 1 RKKEKVRKSKATEFIDYSIE 21

RESULT 7
ADS41968
ID ADS41968 standard; protein; 371 AA.

XX ADS41968;
XX 02-DEC-2004 (first entry)
XX Bacterial polypeptide #20398.
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX Bacteria.
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 20398; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 371 AA;

Query Match 13.0%; Score 90; DB 8; Length 371;
Best Local Similarity 26.8%; Pred. No. 0.1;
Matches 34; Conservative 23; Mismatches 38; Indels 32; Gaps 7;

QY 13 MHNIHILQIRKLHRL-----SNFP-----RLPGILAPETVLLFCYKVF 53

Db 21 MHNL--AIKLRGHEVGIVTNNRPTCKEELKRYGIELIKIPGIISP-----FLDV 70
QY 54 RKKEKVRKQKATEFI-DYSIEQSHAILTPLOTHLTMKGSSM-KCSSLSSEAILFTLTL 111
Db 71 NLTYGLKSSSEELNEFLKDPDIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 129
QY 112 QLTQTILG 118
Db 130 KLWDTILG 136

RESULT 8
ADN18681
ID ADN18681 standard; protein; 381 AA.
AC ADN18681;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #1334.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 1334; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or

CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 381 AA;

Query Match 13.0%; Score 90; DB 8; Length 381;
Best Local Similarity 26.8%; Pred. No. 0.11;
Matches 34; Conservative 23; Mismatches 38; Indels 32; Gaps 7;
QY 13 MNIIHILOIRKLHRL-----SNFP-----RLPGILAPETVLLPFCYKVF 53
Db 24 MHNL--AIKLRGHEVGIVTNNRPTCKEELKRYGIELIKIPGIISP-----FLDV 73
QY 54 RKKEKVRKQKATEFI-DYSIEQSHAILTPLOTHLTMKGSSM-KCSSLSSEAILFTLTL 111
Db 74 NLTYGLKSSSEELNEFLKDPDIHSHHA-FTPLSLKALKAGKNMEKGTLLTTHSISFAHES 132
QY 112 QLTQTILG 118
Db 133 KLWDTILG 139

RESULT 9
ABB90843
ID ABB90843 standard; protein; 645 AA.
XX
AC ABB90843;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 54.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP009892.
XX
PR 28-AUG-2001; 2001WO-EP009892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidner M;
XX WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.
XX
PS Claim 5; SEQ ID NO 54; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX
SQ Sequence 645 AA;

Query Match 11.9%; Score 82.5; DB 5; Length 645;
Best Local Similarity 23.5%; Pred. No. 1.7;
Matches 31; Conservative 26; Mismatches 46; Indels 29; Gaps 5;
QY 25 LHRLSNF---PRLPGILAPETV-----LLPFCYKVRKKEKVKRQ 63
DB 99 MTHLLAGTYGPHWPVMTAVTFKMLTGIVSFLTALSIVTLPLLLKA--KVREFMLSK 156
QY 64 KATEFDYSGSHHAILPLOTHTLTKGSSMKC--SSLSEAILFTLTQLTQTGLECC 122
DB 157 KTRF-----LDREVGIIMQTELSLHVRMLTKIRTSRHTLYTLVELSKTLGKNC 211
QY 123 LLYLSKTIHPQI 134
DB 212 AVWIPNEIKTEM 223
RESULT 10
ADN48099
ID ADN48099 standard; protein; 384 AA.
AC ADN48099;
XX
DT 01-JUL-2004 (first entry)
XX
DE Thermococcus kodakaraensis KOD1 protein sequence SeqID1977.
XX
KW gene disruption; gene targeting; marker gene; transformation;
KW homologous recombination; hyperthermostable archaeobacterium; KOD1;
KW gene structure; gene function; enzyme activity; medicine;
KW forensic science; food; drug inspection; molecular biology; immunology.
XX
OS Thermococcus kodakaraensis.
XX
PN WO2004022736-A1.
XX
PD 18-MAR-2004.
XX
PF 29-AUG-2003; 2003WO-IB003597.
XX
PR 30-AUG-2002; 2002JP-00319011.
XX
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI Imanaka T, Atomi H;
XX
PS WPI; 2004-257583/24.
XX
PT Method for disrupting targeted gene in genome of organism particularly
PT thermostable bacterium and with genome chips for analysis, applicable in
PT studying gene structure and functions.
XX
PS Claim 9; SEQ ID NO 1977; 598pp; Japanese.
XX
CC This invention relates to a novel method for targeting disruption of an
CC arbitrary gene in a genome of an organism which comprises providing the
CC whole sequential data of the genome of such organism, selecting at least
CC 1 arbitrary region in the sequence, providing a vector that contains a
CC sequence homologous with the selected region and a marker gene,
CC transformation, and homologous recombination. The genome is preferably
CC the genome of a hyperthermostable archaeobacterium, particularly
CC Thermococcus kodakaraensis KOD1. The method is for targeting the
CC disruption of a gene in the genome of an organism, which is applicable in
CC studying gene structure and functions as well as enzyme activities of
CC encoded proteins and useful in medicine, forensic science, food or drug
CC inspection, molecular biology and immunology. With this method, the
CC disruption of a gene at an arbitrary position in a genome can be achieved
CC efficiently and reliably. The present sequence is that of a protein
CC encoded by the genome of Thermococcus kodakaraensis which was derived
CC using the method of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 384 AA;
Query Match 11.7%; Score 81; DB 8; Length 384;
Best Local Similarity 23.3%; Pred. No. 1.3;
Matches 31; Conservative 28; Mismatches 42; Indels 32; Gaps 7;
QY 13 MHNIHILQIRKLHRLS-----NFPRLPGILAPETVLLPFCYKVF 53
DB 21 MHQL--AIVLKLGHDSIVNDLKTGKEKELEELGVLGVGVISP-VGINITY--- 74
QY 54 RKKEKVKSSQKATEF-IDYSIEQSHHAILTPLQHTLTKGSSM-KCSLSSEAILFTTUL 111
DB 75 ----GLKSNRELGEFLVDFVVHAHA-FTPLSLKAVKAGRTLEKATLLTHSISFSHES 129
QY 112 QLTQTILGLECCLL 124
DB 130 SLWKALGLTFPLL 142
RESULT 11
ADF50279
ID ADF50279 standard; protein; 1061 AA.
XX
AC ADF50279;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human PFM7 protein, a PR/SET family member.
XX
KW human; PFM7; SET; Su(var)3-9, Enhancer-of-Zeste and Trithorax; cancer;
KW PFM/SET; tumour suppressor; hyperproliferative disorder; scleroderma;
KW arthritis; alcoholic liver cirrhosis; hypertrophic scarring;
KW atherosclerosis; gene therapy; Duchenne's muscular dystrophy;
KW insulin-dependent diabetes mellitus; Huntington's; Parkinson's;
KW Alzheimer's disease; paralysis; cerebellar atrophy; cytostatic;
KW neuroprotective; nootropic; antiarthritic; vulnerary;
KW antiatherosclerotic; antidiabetic; hepatotropic; PFM; PR family member.
XX
OS Homo sapiens.
XX
PN US2003049623-A1.
XX
PD 13-MAR-2003.
XX
PF 18-JUL-2001; 2001US-00910478.
XX
PR 18-JUL-2001; 2001US-00910478.
XX
PA (HUAN/) HUANG S.
XX
PI Huang S;
XX
PS WPI; 2003-567062/53.
XX
DR N-PSDB; ADF50278.
XX
PT New PR/SET-domain containing nucleic acids (which encodes PFM/SET) and
PT polypeptides, useful for preventing or treating cancers, scleroderma,
PT arthritis, keloids, atherosclerosis, Huntington's disease or Alzheimer's
PT disease.
XX
PS Claim 21; SEQ ID NO 4; 83pp; English.
XX
CC This invention relates to novel PFM (PR family member)/SET (Su(var)3-9,
CC Enhancer-of-Zeste and Trithorax) nucleotides and encoded polypeptides
CC thereof that are implicated in proliferative disorders such as cancer.
CC Specifically, it refers to a functional fragment of the PFM/SET tumour
CC suppressor gene that encodes a PR, SET, PRAZ or PKZL domain. The present
CC invention describes the identification and characterisation of additional
CC PR/SET-domain family members that can be used as regulators of cell
CC proliferation and furthermore to treat, prevent or diagnose
CC hyperproliferative disorders including scleroderma, arthritis, alcoholic

CC liver cirrhosis, hypertrophic scarring and atherosclerosis. Through gene
CC therapy, these polynucleotides can be used to enhance proliferation of
CC normal cells without rendering the cells cancerous and as such they are
CC particularly useful for treating Duchenne's muscular dystrophy, insulin-
CC dependent diabetes mellitus, Huntington's, Parkinson's, Alzheimer's
CC disease, paralysis, or cerebellar atrophy. Accordingly, these
CC compositions can be described as cytostatic, neuroprotective, nontropic,
CC antiarthritic, vulnary, antiatherosclerotic, antidiabetic or
CC hepatotropic. This polypeptide sequence is the human PFM7 protein
CC (encoded by a PFM/SET gene localised to chromosome 11q25) of the
CC invention.
XX
SQ Sequence 1061 AA;

Query Match 11.2%; Score 77.5; DB 7; Length 1061;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 36; Conservative 17; Mismatches 42; Indels 31; Gaps 6;

QY 18 HILQ-----IRKLR-----HRLSNFRLPGILAPETVLLPFCYKVRKKEKV 59
Db 704 HILKNHPGAELPPSIRKLRPAGGEPDMLSTHTLTGTIATPPVCCPHCSKQYSSKTKM 763

QY 60 KR--SQKATEFDYSIEQSHHAILTLOTH-LTMKGSSMKCSLSSEAILFTLTL----- 111
Db 764 VQHIRKHPFAQLS-----NTIHTEPLTAVISATPAVLTTDSATGETVVTDLTLTQAMT 818

QY 112 QLTQTL 117
Db 819 ELSQTL 824

RESULT 12
ID ADO36741
AC ADO36741; standard; protein; 1061 AA.
XX
XX ADO36741;
DT 15-JUL-2004 (first entry)
DE Human PR-domain containing protein PFM-7.
KW Cytostatic; histone methyltransferase; gene therapy;
KW PFM/SET protein binding motif; cell growth modulator;
KW histone methyltransferase activity; proliferative disorder; cancer;
KW PR-domain; PFM-7.
XX
OS Homo sapiens.
XX
XX US2004014192-A1.
XX
XX 22-JAN-2004.
XX
XX 18-JUL-2002; 2002US-00200012.
XX
XX 18-JUL-2002; 2002US-00200012.
XX
XX (HUAN/) HUANG S.
XX
XX Huang S;
XX
XX WPI; 2004-121568/12.
XX
XX
XX New isolated nucleic acid molecule comprising a sequence encoding a
XX PFM/SET polypeptide, useful for diagnosing, prognosing, preventing and
XX treating proliferative disorders, e.g. cancer.
XX
XX Claim 1; SEQ ID NO 4; 38pp; English.
XX
XX The invention describes an isolated nucleic acid molecule comprising a
XX sequence encoding a PFM/SET protein binding motifs polypeptides
XX comprising a 689, 1061, 367, 717, 571, 951, 1257, 720, or 770 amino acid
XX sequence (p1-p9), given in the specification. Also described are: a
XX vector comprising the isolated nucleic acid molecule; a host cell

CC comprising the vector; an oligonucleotide comprising at least 17
CC contiguous nucleotides of the nucleic acid molecule; a primer pair for
CC detecting PFM/SET nucleic acid molecule, comprising two isolated
CC oligonucleotides; detecting PFM/SET nucleic acid molecule in a sample;
CC modulating cell growth by introducing the vector into a host cell, and
CC expressing the encoded PFM/SET polypeptide in an amount effective to
CC modulate growth of the cell; isolating PFM/SET polypeptide by growing the
CC host cell under conditions appropriate for the expression of the
CC polypeptide; the isolated PFM/SET polypeptide, or its functional fragment
CC; an isolated immunogenic PFM/SET peptide, comprising at least 8
CC contiguous amino acids of p1-p9; an antibody or its antigen-binding
CC fragment that specifically binds to PFM/SET polypeptide; detecting
CC PFM/SET polypeptide in a sample; and screening for a compound that
CC modulates PFM/SET polypeptide histone methyltransferase activity by
CC contacting the polypeptide with one or more candidate compounds, and
CC determining histone methyltransferase activity of the contacted
CC polypeptide. The nucleic acid molecules, polypeptides and methods are
CC useful for diagnosing, prognosing, preventing and treating proliferative
CC disorders, e.g. cancer. This is the amino acid sequence of PR-domain
CC containing polypeptide PFM-7.
XX
SQ Sequence 1061 AA;

Query Match 11.2%; Score 77.5; DB 8; Length 1061;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 36; Conservative 17; Mismatches 42; Indels 31; Gaps 6;

QY 18 HILQ-----IRKLR-----HRLSNFRLPGILAPETVLLPFCYKVRKKEKV 59
Db 704 HILKNHPGAELPPSIRKLRPAGGEPDMLSTHTLTGTIATPPVCCPHCSKQYSSKTKM 763

QY 60 KR--SQKATEFDYSIEQSHHAILTLOTH-LTMKGSSMKCSLSSEAILFTLTL----- 111
Db 764 VQHIRKHPFAQLS-----NTIHTEPLTAVISATPAVLTTDSATGETVVTDLTLTQAMT 818

QY 112 QLTQTL 117
Db 819 ELSQTL 824

RESULT 13
ABBS8124
ID ABBS8124 standard; protein; 1144 AA.
XX
XX ABBS8124;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 1164.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX N-PSDB; ABL02227.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
PT

PT interactions.
XX Disclosure; SEQ ID NO 1164; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1144 AA;

Query Match 11.2%; Score 77.5; DB 4; Length 1144;
Best Local Similarity 30.0%; Pred. No. 15;
Matches 30; Conservative 12; Mismatches 33; Indels 25; Gaps 5;

QY 15 NIHIHQIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKK-----EKVRSQKATEF 68
DB 718 NQVHEIQ-RLARHVSNNL-----KAIDPTNFRFFFLKDRHGQEKAK--SOIQKF 765

QY 69 IDYSIBQSH-----HAILTLQTLTKGSSMKCSSLS 101
DB 766 LNFIEDDHNGSEAIYTFLSPSDHLKQSLPSPKSKFS 805

RESULT 14
ADS08184
ID ADS08184 standard; protein; 1026 AA.
XX
AC ADS08184;
XX
DT 04-NOV-2004 (first entry)
DE Staphylococcus epidermis polypeptide seqid 7479.
XX
KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW recombinant expression vector; infection; computer readable medium;
KW computer based system.
XX
OS Staphylococcus epidermidis.
XX
PN US2004147734-A1.
XX
PD 29-JUL-2004.
XX
PF 01-DEC-2003; 2003US-00724972.
XX
PR 08-NOV-1997; 97US-0064964P.
PR 13-AUG-1998; 98US-00134001.
PR 29-NOV-1999; 99US-00450969.
XX
PA (DOUC/) DOUCETTE-STAMM L.
PA (BUSH/) BUSH D.
XX
PI Doucette-Stamm L, Bush D;
XX
DR WPI; 2004-580138/56.
DR N-PSDB; ADS04412.
XX
PT New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.
XX
PS Claim 17; SEQ ID NO 7479; 741pp; English.
XX
CC The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any

CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC given in the specification. Also described are: a recombinant expression
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an S. epidermidis polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for S. epidermidis infection; a recombinant or substantially
CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
CC infection. This is the amino acid sequence of a S. epidermis protein of
CC the invention.
XX
SQ Sequence 1026 AA;

Query Match 11.1%; Score 77; DB 8; Length 1026;
Best Local Similarity 20.6%; Pred. No. 15;
Matches 35; Conservative 33; Mismatches 54; Indels 48; Gaps 8;

QY 7 NQKYED-MNIIHILOIR-----KLRHLSNFPRLPGILAPETVL----LPFCYKV 52
DB 782 NQALEDEMHTVIKIVELSRQARKNADLKIKQPLSKM-----VIRPNSQLNLSFLPNYYSI 836

QY 53 FRKEKVR---SQKATEFIDYSIEQSHAI-----LTPLQTLTKWG 92
DB 837 IKDELNIKNIELTDINDYITVELKNFNSVGPKNLGNKTKNIQTLIDLSLEYDKSLIES 896

QY 93 SSMKCSSLSSEA-----ILFTLTQLTQTLGLECCLLYLSKTHPOII 135
DB 897 NNFK--SUSSDAELTKDFFIKTLPKDSYQLSENDCVILLDKNLSPELI 944

RESULT 15
ABM69043
ID ABM69043 standard; protein; 322 AA.
XX
AC ABM69043;
XX
DT 20-NOV-2003 (first entry)
DE Photorhabdus luminescens protein sequence #2140.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PS (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Duchaud B, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
Buchrieser C;

XX WPI; 2003-148459/14.
DR
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX
PS Claim 2; SEQ ID NO 2140; 1205pp; French.
PS
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 322 AA;
Query Match 11.0%; Score 76.5; DB 6; Length 322;
Best Local Similarity 26.6%; Pred. No. 3.5;
Matches 29; Conservative 19; Mismatches 50; Indels 11; Gaps 3;
QY 13 MNNIHILQIRKLRLSNFPRPLPGILAPETVLLPFCYKVKRKKRSOKATEPIDYS 72
Db 221 IHNFIHAVQIPAEKGLSRVTNLPFGISVTVQEMIDALAEV-----AGQKAVELIRFE 272
QY 73 IEQSHHAILTPLOTHLM-KGSSM--KCSLSSEAILFTTLTQLTQLG 118
Db 273 PDENINRIVASWFGHFDISRGLSLGFHADNTFTDTRAFITNNLSQTGG 321

Search completed: September 9, 2005, 13:00:52
Job time : 59.5 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:25:33 ; Search time 17.5 Seconds
(without alignments)
575.864 Million cell updates/sec

Title: US-09-974-546C-86

Perfect score: 694

Sequence: 1 MRAFLRNQKYEDMHNIIHL.....TLGLECLLYLSKTIHPQII 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/prodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/PCUTS_COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	135	3	US-09-097-199-84
2	694	100.0	135	3	US-09-097-199-86
3	694	100.0	157	4	US-09-949-016-9182
4	103	14.8	21	3	US-09-097-199-56
5	72	10.4	636	4	US-09-614-912-92
6	71.5	10.3	634	4	US-09-614-912-88
7	70.5	10.2	438	4	US-09-489-039A-11150
8	70.5	10.2	634	4	US-09-614-912-100
9	69.5	10.0	1294	3	US-09-473-717-2
10	69.5	10.0	1294	4	US-09-949-016-6023
11	69.5	10.0	1353	3	US-09-398-193-99
12	68.5	9.9	2470	4	US-08-265-967C-2
13	68.5	9.9	2470	4	US-08-305-790B-3
14	68	9.8	496	4	US-10-146-704-3
15	68	9.8	639	4	US-09-949-016-6812
16	68	9.8	652	4	US-09-949-016-7323
17	68	9.8	993	4	US-09-949-016-10335
18	67.5	9.7	156	4	US-09-543-681A-7434
19	67	9.7	399	4	US-09-543-681A-4933
20	67	9.7	517	3	US-09-457-040B-14
21	67	9.7	773	2	US-08-484-101B-42
22	67	9.7	773	2	US-08-484-101B-44
23	67	9.7	773	3	US-08-714-524D-42
24	67	9.7	773	3	US-08-714-524D-44
25	66.5	9.6	580	4	US-09-107-532A-5734
26	66.5	9.6	1784	3	US-09-040-738-2
27	66.5	9.6	1784	3	US-08-652-426A-2

28	66.5	9.6	1813	4	US-09-949-016-8283	Sequence 8283, Ap
29	66.5	9.6	1813	4	US-09-949-016-8284	Sequence 8284, Ap
30	66.5	9.6	1813	4	US-09-949-016-8285	Sequence 8285, Ap
31	66	9.5	267	4	US-09-270-767-46386	Sequence 46386, A
32	66	9.5	348	4	US-09-198-452A-922	Sequence 922, App
33	66	9.5	348	4	US-09-438-185A-856	Sequence 856, App
34	66	9.5	611	4	US-09-949-016-7139	Sequence 7139, App
35	65.5	9.4	698	4	US-09-538-092-151	Sequence 151, App
36	65.5	9.4	3218	1	US-08-764-100-27	Sequence 27, Appl
37	65	9.4	75	4	US-09-621-976-6155	Sequence 6155, Ap
38	65	9.4	258	4	US-09-270-767-33067	Sequence 33067, A
39	65	9.4	258	4	US-09-270-767-48284	Sequence 48284, A
40	65	9.4	288	1	US-08-375-709-9	Sequence 9, Appli
41	65	9.4	288	1	US-08-752-923-9	Sequence 9, Appli
42	65	9.4	288	3	US-09-090-793-6	Sequence 6, Appli
43	65	9.4	288	4	US-09-231-899-6	Sequence 6, Appli
44	64.5	9.3	236	4	US-09-107-532A-6873	Sequence 6873, Ap
45	64.5	9.3	315	3	US-09-134-001C-5446	Sequence 5446, Ap

ALIGNMENTS

RESULT 1

US-09-097-199-84
; Sequence 84, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Veltri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-097-199-84

Query Match 100.0%; Score 694; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 4.7e-77;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEVK 60
DB 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEVK 60
QY 61 RSQKATEFIDYIEQSHHAILPQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTLGLE 120
DB 61 RSQKATEFIDYIEQSHHAILPQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTLGLE 120
QY 121 CCLLYLSKTIHPQII 135
DB 121 CCLLYLSKTIHPQII 135
RESULT 2
US-09-097-199-86
; Sequence 86, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Velttri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/692,787
; FILING DATE: 31-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakaehima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-097-199-86
Query Match 100.0%; Score 694; DB 3; Length 135;
Best Local Similarity 100.0%; Pred. No. 4.7e-77;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEVK 60
DB 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEVK 60
QY 61 RSQKATEFIDYIEQSHHAILPQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTLGLE 120
DB 61 RSQKATEFIDYIEQSHHAILPQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTLGLE 120
QY 121 CCLLYLSKTIHPQII 135
DB 121 CCLLYLSKTIHPQII 135

DB 121 CCLLYLSKTIHPQII 135
RESULT 3
US-09-949-016-9182
; Sequence 9182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9182
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9182
Query Match 100.0%; Score 694; DB 4; Length 157;
Best Local Similarity 100.0%; Pred. No. 5.8e-77;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEVK 60
DB 23 MRAFLRNQKYEDMHNIIHILOIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEVK 82
QY 61 RSQKATEFIDYIEQSHHAILPQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTLGLE 120
DB 83 RSQKATEFIDYIEQSHHAILPQTHLTMKGSSMKCSSLSSEAILFTLTQLTQTLGLE 142
QY 121 CCLLYLSKTIHPQII 135
DB 143 CCLLYLSKTIHPQII 157
RESULT 4
US-09-097-199-56
; Sequence 56, Application US/09097199
; Patent No. 6218529
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; APPLICANT: O'Hara, S. Mark
; APPLICANT: Ralph, David
; APPLICANT: Velttri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; TITLE OF INVENTION: PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,199
; FILING DATE:
; CLASSIFICATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 12:59:09 ; Search time 52.5 Seconds
(without alignments)
1014.252 Million cell updates/sec

Title: US-09-974-546C-86

Perfect score: 694

Sequence: 1 MRAFLRNQKVEDMHNIIHL.....TLGLECLLYLSKTHPQII 135

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1777461 seqs, 394431504 residues

Total number of hits satisfying chosen parameters: 1777461

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	135	10	US-09-974-546-84
2	694	100.0	135	10	US-09-974-546-86
3	103	14.8	21	10	US-09-974-546-56
4	90	13.0	371	15	US-10-369-493-20398
5	90	13.0	381	15	US-10-369-493-1334
6	81	11.7	1478	16	US-10-425-115-362997
7	79	11.4	264	15	US-10-424-599-174106
8	79	11.4	1368	16	US-10-425-115-324108
9	79	11.4	1444	16	US-10-425-115-324095
10	78	11.2	289	15	US-10-424-599-148597
11	78	11.2	3229	16	US-10-425-115-324093

12 77.5 11.2 1061 15 US-10-200-012-4
13 77 11.1 1026 18 US-10-724-972A-7479
14 77 11.1 1766 16 US-10-425-115-324098
15 633 15 US-10-282-122A-70203
16 75 10.8 1259 16 US-10-425-115-324122
17 75 10.8 3749 16 US-10-425-115-324125
18 74.5 10.7 388 15 US-10-424-599-169767
19 74.5 10.7 404 18 US-10-954-778-97
20 74 10.7 1353 16 US-10-425-115-362744
21 74 10.7 1605 16 US-10-425-115-362998
22 74 10.7 1676 16 US-10-425-115-238569
23 74 10.7 1708 16 US-10-425-115-324103
24 74 10.7 1752 16 US-10-425-115-238562
25 74 10.7 1754 16 US-10-425-115-324104
26 73 10.5 1143 16 US-10-425-115-238564
27 73 10.5 1517 16 US-10-425-115-324099
28 72.5 10.4 3365 16 US-10-739-930-6537
29 72 10.4 329 15 US-10-424-599-150082
30 71.5 10.3 407 15 US-10-425-114-69882
31 71.5 10.3 426 16 US-10-425-115-327356
32 71.5 10.3 676 15 US-10-424-599-267862
33 71 10.2 334 15 US-10-425-114-63742
34 71 10.2 1214 16 US-10-425-115-243074
35 71 10.2 1595 16 US-10-425-115-238559
36 71 10.2 1715 16 US-10-425-115-324123
37 71 10.2 1924 9 US-09-866-557A-2
38 71 10.2 1924 11 US-09-858-862-2
39 71 10.2 1924 14 US-10-055-797-2
40 71 10.2 1924 15 US-10-350-798-2
41 70.5 10.2 269 15 US-10-424-599-264790
42 70.5 10.2 634 17 US-10-876-086-27
43 70 10.1 206 16 US-10-408-765A-1698
44 70 10.1 272 16 US-10-437-963-125982
45 70 10.1 327 15 US-10-425-114-63743

ALIGNMENTS

RESULT 1

US-09-974-546-84

; Sequence 84, Application US/09974546

; Publication No. US20030050470A1

; GENERAL INFORMATION:

; APPLICANT: An, Gang

; O'Hara, S. Mark

; Ralph, David

; Veltri, Robert

; TITLE OF INVENTION: BIONKERS AND TARGETS FOR DIAGNOSIS,

; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE

; NUMBER OF SEQUENCES: 87

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: USA

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION NUMBER: US/09/974,546

; FILING DATE: 10-Oct-2001

; CLASSIFICATION: Unknown

; PRIORITY DATA:

; APPLICATION NUMBER: 09/097,199

; FILING DATE: 1998-06-12

; ATTORNEY/AGENT INFORMATION:

; NAME: Nakashima, Richard A.

; REGISTRATION NUMBER: P-42,023

```

; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-09-974-546-84

Query Match          100.0%; Score 694; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 9.1e-72;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLRNQKYEDMHNIIHILQIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEKVK 60
Db 1 MRAFLRNQKYEDMHNIIHILQIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEKVK 60

QY 61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
Db 61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120

QY 121 CCLLYLSKTIHPQII 135
Db 121 CCLLYLSKTIHPQII 135

RESULT 2
US-09-974-546-86
; Sequence 86, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Velttri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/974,546
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/097,199
; FILING DATE: 1998-06-12
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 86:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 135 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

Query Match          100.0%; Score 694; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 9.1e-72;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLRNQKYEDMHNIIHILQIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEKVK 60
Db 1 MRAFLRNQKYEDMHNIIHILQIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEKVK 60

QY 61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
Db 61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120

QY 121 CCLLYLSKTIHPQII 135
Db 121 CCLLYLSKTIHPQII 135
```

```

; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 86:
US-09-974-546-86

Query Match          100.0%; Score 694; DB 10; Length 135;
Best Local Similarity 100.0%; Pred. No. 9.1e-72;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRAFLRNQKYEDMHNIIHILQIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEKVK 60
Db 1 MRAFLRNQKYEDMHNIIHILQIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEKVK 60

QY 61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120
Db 61 RSQKATEFIDYSIEQSHHAILTPLQTHLTMKGSSMKCSSLSSEAILFTLTQLTOTLGL 120

QY 121 CCLLYLSKTIHPQII 135
Db 121 CCLLYLSKTIHPQII 135

RESULT 3
US-09-974-546-56
; Sequence 56, Application US/09974546
; Publication No. US20030050470A1
; GENERAL INFORMATION:
; APPLICANT: An, Gang
; O'Hara, S. Mark
; Ralph, David
; Velttri, Robert
; TITLE OF INVENTION: BIOMARKERS AND TARGETS FOR DIAGNOSIS,
; PROGNOSIS AND MANAGEMENT OF PROSTATE DISEASE
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/974,546
; FILING DATE: 10-Oct-2001
; CLASSIFICATION: Unknown
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/097,199
; FILING DATE: 1998-06-12
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakashima, Richard A.
; REGISTRATION NUMBER: P-42,023
; REFERENCE/DOCKET NUMBER: UROC:018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-974-546-56

Query Match          14.8%; Score 103; DB 10; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 RKKEKVKRSQKATEFIDYSIE 74
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 13:01:39 ; Search time 14.5 Seconds
(without alignments)
895.811 Million cell updates/sec

Title: US-09-974-546C-86

Perfect score: 694

Sequence: 1 MRAFLRNQKYEDMHNIIHL.....TLGLECLLYLSTKTHPQII 135

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 79:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	13.0	381	2 F71196	probable hexosyltr
2	82.5	11.9	645	2 F86174	hypothetical prote
3	80	11.5	495	1 MNXBSA	nonstructural prot
4	78.5	11.3	3147	2 T21328	hypothetical prote
5	78	11.2	523	2 D64555	conserved hypothet
6	74.5	10.7	172	2 A58456	X protein - human
7	74.5	10.7	263	2 A82069	hypothetical prote
8	72.5	10.4	796	2 T43782	hypothetical prote
9	71.5	10.3	227	2 B84040	ABC transporter (A
10	71.5	10.3	451	2 D70045	two-component sens
11	70.5	10.2	1084	2 T40751	isoleucyl-trna syn
12	70	10.1	404	2 T06761	hypothetical prote
13	70	10.1	982	1 A44831	phosphoenolpyruvat
14	70	10.1	982	2 AE2413	phosphoenolpyruvat
15	69.5	10.0	1398	2 H71606	hypothetical prote
16	69	9.9	154	2 T27983	hypothetical prote
17	69	9.9	743	2 T00634	hypothetical prote
18	69	9.9	1850	2 AC1917	serine/threonine k
19	68.5	9.9	503	1 S11338	steroid 11beta-mon
20	68.5	9.9	2470	2 S57085	1-phosphatidylinos
21	68	9.8	379	2 C75006	hypothetical prote
22	68	9.8	414	1 H64203	histidine-tRNA lig
23	68	9.8	496	2 A31986	glucose transporte
24	68	9.8	639	2 T13151	adaptor protein CM
25	68	9.8	5126	2 S40450	ryanodine receptor
26	67.5	9.7	619	2 B87682	hypothetical prote
27	67.5	9.7	833	2 T28385	ORF MSV224 probabl
28	67.5	9.7	1278	2 A71609	probable secreted
29	67.5	9.7	3225	2 D81702	adherence factor T

30	67	9.7	126	2 T03757	hypothetical prote
31	67	9.7	213	2 E96506	hypothetical prote
32	67	9.7	317	2 B82440	hypothetical prote
33	67	9.7	347	2 T01452	translation initia
34	67	9.7	517	2 A54099	protein kinase Dar
35	67	9.7	587	2 S37927	hypothetical prote
36	66.5	9.6	244	2 AC3341	pricornin-2 C20-me
37	66.5	9.6	447	2 T21716	hypothetical prote
38	66.5	9.6	635	2 T06537	ethylene receptor
39	66.5	9.6	859	2 H70327	DNA mismatch repa
40	66.5	9.6	908	2 T22376	hypothetical prote
41	66.5	9.6	1784	2 A49420	tuberosus sclerosis
42	66	9.5	344	2 D86597	outer membrane pro
43	66	9.5	344	2 H72027	major outer membra
44	66	9.5	446	2 A42029	transcription fact
45	66	9.5	472	2 F97147	probable Fe-S oxid

ALIGNMENTS

RESULT 1

F71196
Probable hexosyltransferase (EC 2.4.1.1-) PH1844 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: F71196
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic au
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: F71196
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-381 <KAW>
A:Cross-references: UNIPROT:O59512; GB:AP000007; NID:G3236134; PIDN:BAA30965.1; PID:G3258
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1844
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 13.0%; Score 90; DB 2; Length 381;

Best Local Similarity 26.8%; Pred.No. 0.16;

Matches 34; Conservative 23; Mismatches 38; Indels 32; Gaps 7;

QY 13 MHNIIHILQIRKLRLHRL-----SNFP-----RLPGILAPETVLLPFCYKVF 53

Db 24 MHNLL--AIKLRRRGHEVGIVTNNRPTGKEELKRYGIELIKIPGIISP-----FLDV 73

QY 54 RKKEKVKRSQKATEFI-DYSIEQSHHAILTPLOTHLTKGSSM-KCSSLSBAILFTLTL 111

Db 74 NLTYGLKSSSEELNEFLKOFDIHSHHA-FTPLSLKALKAGKNMGKTLTTHSISFAHES 132

QY 112 QLTQTIG 118

Db 133 KLWDTLIG 139

RESULT 2

F86174

Hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C:Accession: F86174

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ket, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: F86174
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-645 <STO>
A;Cross-references: UNIPROT:P93825; GB:AE005172; NID:g2341041; PIDN:AAB70445.1; GSPDB:GN00023; CESP:F25C8

C;Genetics:
A;Map position: 1
Query Match 11.9%; Score 82.5; DB 2; Length 645;
Best Local Similarity 23.5%; Pred. No. 1.7;
Matches 31; Conservative 26; Mismatches 46; Indels 29; Gaps 5;
QY 25 LHRLSNF---PRLPGILAPETV-----LLPCYKVRKKEKVRQ 63
Db 99 MTHLLAGFTYGHWPVWMTAVTFKMLTGIVSFLTALS LVTLPLLLKA--KYREFMLSK 156
QY 64 KATEFIDYIEQSHHAILTPLOTHLTMKGSSMKC--SSLSEAILFTLTQLTQTLGLECC 122
Db 157 KYRE-----LDREVGIIMQTEISLHVRMLTKIRTSDRHTILYTLVELSKTLGKNC 211
QY 123 LLYLSKTIHPQI 134
Db 212 AVNIPNEIKTEM 223

RESULT 3
MNXRSA
nonstructural protein NCPV2 - simian rotavirus SA11
N;Alternate names: nonstructural protein NS53
C;Species: simian rotavirus SA11
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
C;Accession: S08215
R;Mitchell, D.B.; Both, G.W.
Virolgy 174, 618-621, 1990
A;Title: Conservation of a potential metal binding motif despite extensive sequence divergence
A;Reference number: S08215; MUID:90163231; PMID:2154894
A;Accession: S08215
A;Molecule type: Genomic RNA
A;Residues: 1-495 <MIT>
A;Cross-references: UNIPROT:P15687; EMBL:X14914; NID:g61889; PIDN:CAA33039.1; PID:g61890
C;Genetics:
A;Map position: segment 5
C;Superfamily: bovine rotavirus nonstructural protein NCPV2
C;Keywords: nonstructural protein; zinc finger

Query Match 11.5%; Score 80; DB 1; Length 495;
Best Local Similarity 25.3%; Pred. No. 2.3;
Matches 24; Conservative 22; Mismatches 33; Indels 16; Gaps 4;
QY 2 RAFLRNQKEDMHNIIHILOIRKLRHLSNFPRLPGILAPETVLLPFCYKVRKKEKVR 61
Db 71 RCFLDNEP-----HLKRLTRVHPITK-DKLCQIIDLYNIIFPINDKVRKFERMIK 121
QY 62 SOKATEFIDYIEQSHHAILTPLOTHLTMKGSSMK 96
Db 122 QRKCRN--QYKIEWNHLLP-----ITLNAARF 149

RESULT 4
T21328
hypothetical protein F25C8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T21328
R;Matthews, L.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19407
A;Accession: T21328
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA
A;Residues: 1-3147 <WIL>
A;Cross-references: UNIPROT:Q9XV66; EMBL:Z81512; PIDN:CAB04172.1; GSPDB:GN00023; CESP:F25C8
A;Experimental source: clone F25C8
C;Genetics:
A;Gene: CESP:F25C8.3
A;Map position: 5
A;Introns: 32/2; 57/2; 73/3; 113/3; 157/3; 366/1; 456/1; 509/1; 638/1; 773/2; 848/1; 902/3; 2510/2; 2737/3; 2827/1; 2902/3; 3093/3; 3121/1
C;Superfamily: Caenorhabditis elegans hypothetical protein F25C8.3

Query Match 11.3%; Score 78.5; DB 2; Length 3147;
Best Local Similarity 28.1%; Pred. No. 26;
Matches 41; Conservative 17; Mismatches 67; Indels 21; Gaps 6;
QY 2 RAFLRN-----QKVEDMHNIIHILOIRKLRHLSNFPRLPGILAPETVLLPFCYKVR 54
Db 1571 RADKENLADNMQAKQALRKSIHARQSTAVPRRESAMVGQPE-PASKAIKMLMEKMQQ 1629
QY 55 KKEKVRKSKATE--FIDYIEQSHHAILTPLOTHLTMKGSSMKCSSLSSEAILFTLTQ 112
Db 1630 EKEKEKEKEKEDALKKQSQVEQDSSTDEDAQLPEKNPML-----TYLRSVLQ 1682
QY 113 ITQ---TLGLECCLLYLSKTIHPQII 135
Db 1683 LVHSPISSVLVKCCLL-LSVEQHKQMI 1707

RESULT 5
D64555
conserved hypothetical integral membrane protein HP0284 - Helicobacter pylori (strain 266
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: D64555
R;Tombs, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.N
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: D64555
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-523 <TOM>
A;Cross-references: UNIPROT:Q25059; GB:AE000547; GB:AE000511; NID:g2313377; PIDN:AAD07352

Query Match 11.2%; Score 78; DB 2; Length 523;
Best Local Similarity 23.5%; Pred. No. 3.9;
Matches 24; Conservative 16; Mismatches 30; Indels 32; Gaps 3;
QY 5 LRNQKVEDMHNIIHILOIRKLRHLS-----NPPRLPGILA 40
Db 71 LKNKSKD-----LLKISTLEHTLKALESQOKMPESYGVNPFKDLIERPNIPNIPN 124
QY 41 PETVL--LPFCYKVRKKEKVRKSKATEFIDYIEQSHHAI 80
Db 125 PIAIIDGISFIKSMHLKHENLRNQTALVEVRLDDQKHQL 166

RESULT 6
A58456
X protein - human hepatitis B virus
C;Species: human hepatitis B virus
C;Date: 30-Sep-1993 #sequence_revision 21-Jan-1997 #text_change 09-Jul-2004
C;Accession: A58456
R;Zeng, M.; Huang, B.R.; Cai, L.W.; Pan, G.Z.
Chinese Biochem. J. 12, 22-26, 1996
A;Title: The sequence analysis of the hepatitis B virus (HBV) X gene and its expression i
A;Reference number: JCL131
A;Accession: A58456
A;Molecule type: DNA
A;Residues: 1-172 <ZEN>

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2005, 13:00:59 ; Search time 51.5 Seconds
(without alignments)
1342.343 Million cell updates/sec

Title: US-09-974-546C-86

Perfect score: 694

Sequence: 1 MRAFLRNQYEDMHNIIHL.....TLGLBCLLYLSKTIHPQII 135

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_03.*

1: uniprot_prot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	694	100.0	135	2	Q9GZY1
2	90	13.0	281	2	O59512
3	82.5	11.9	645	2	P93825
4	81	11.7	451	2	Q7S2X8
5	80	11.5	414	2	Q99FX6
6	80	11.5	479	2	Q99FX4
7	80	11.5	495	1	VNS3
8	80	11.5	496	2	Q99FX5
9	80	11.5	496	2	Q99FX7
10	78.5	11.3	3175	2	Q7JKT8
11	78	11.2	523	2	O25059
12	78	11.2	1646	2	Q86820
13	77.5	11.2	358	2	O8T2U8
14	77.5	11.2	931	2	Q9NKA14
15	77.5	11.2	1023	2	O8G3Z2
16	77.5	11.2	1061	1	PRDA_HUMAN
17	77.5	11.2	1061	2	Q9ULI9
18	77.5	11.2	1117	2	O8T4E8
19	77.5	11.2	1117	2	Q9W3N0
20	77	11.1	495	1	VNS3
21	77	11.1	1024	1	SYIP_STAUA
22	76.5	11.0	321	2	Q7N442
23	76.5	11.0	398	2	Q6LRR8
24	76.5	11.0	2259	1	YCF2_PHYPA
25	75	10.8	629	2	O6GJ08
26	74.5	10.7	160	2	O6SIF8
27	74.5	10.7	163	2	O6SCJ3
28	74.5	10.7	172	2	Q7M022
29	74.5	10.7	263	2	Q9KP79
30	74.5	10.7	404	2	Q7PNX8
31	74	10.7	3184	2	Q9XV66

32	73.5	10.6	615	2	O9FHK4
33	73.5	10.6	853	2	Q7R2I5
34	72.5	10.4	220	2	Q7P537
35	72.5	10.4	220	2	Q8RF89
36	72.5	10.4	716	2	O21040
37	72.5	10.4	796	2	O9TGM3
38	72.5	10.4	1283	2	O9T034
39	72.5	10.4	3306	2	Q9FT44
40	72	10.4	316	2	Q83DP2
41	72	10.4	400	2	Q6RX30
42	72	10.4	448	2	O6ALV1
43	72	10.4	475	2	O8XLZ4
44	72	10.4	788	2	Q7RJY2
45	71.5	10.3	227	2	Q9K884

ALIGNMENTS

RESULT 1
Q9GZY1 PRELIMINARY; PRT; 135 AA.
AC Q9GZY1;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE UC28 protein.
GN Name=UC28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate Cancer;
RX MEDLINE=21028101; PubMed=11156405;
RA An G., Ng A.Y., Meka C.S.R., Luo G., Bright S.P., Cazares L.,
RA Wright G.L. Jr., Veltri R.W.;
RT "Cloning and characterization of UROC28, a novel gene overexpressed in
RT prostate, breast, and bladder cancers.";
RL Cancer Res. 60:7014-7020(2000).
DR EMBL; AF189270; AAG17118.1;
DR EMBL; AF189269; AAG17117.1;
DR Genew; HGNC:21079; PBOV1.
SQ SEQUENCE 135 AA; 15722 MW; 2B7DB8B81983705D0 CRC64;

Query Match 100.0%; Score 694; DB 2; Length 135;
Best Local Similarity 100.0%; Pred.No. 2.9e-60; Mismatches 0; Indels 0; Gaps 0;
Matches 135; Conservative 0;
QY 1 MRAFLRNQYEDMHNIIHLQIRKLRLHLSNFPRLPGILAPETVLLPFCYKVRKKEVKV 60
Db 1 MRAFLRNQYEDMHNIIHLQIRKLRLHLSNFPRLPGILAPETVLLPFCYKVRKKEVKV 60
QY 61 RSQKATEFIDYSIEQSHAILTPLQTHLTMKGSSKCSSLSSEALFTLTLTQTLGLE 120
Db 61 RSQKATEFIDYSIEQSHAILTPLQTHLTMKGSSKCSSLSSEALFTLTLTQTLGLE 120
QY 121 CCLLYLSKTIHPQII 135
Db 121 CCLLYLSKTIHPQII 135

RESULT 2
O59512 PRELIMINARY; PRT; 381 AA.
ID O59512
AC O59512;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein PH1844.
GN OrderedlocusNames=PH1844;
OS Pyrococcus horikoshii.

```

RP SEQUENCE FROM N.A.
RA Theologias;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Hua J., Sakail H., Nourizadeh S., Chen Q.G., Bleecker A.B.,
RL Ecker J.R., Meyerowitz E.M.;
RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Sakai H., Hua J., Chen Q.G., Chang C., Medrano L.J., Bleecker A.B.,
RL Meyerowitz E.M.;
RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AC000104; AAB70445.1; -.
DR EMBL; AF047976; AAC62209.1; -.
DR PIR; F86174; F86174.
DR GO; GO:0015020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR003661; His_kinA_N.
DR InterPro; IPR011052; Prot_amy1_inhib.
DR Pfam; PF01590; GAF; 1.
DR DR PFam; PF00512; HsKA; 1.
DR DR SMART; SM00065; GAF; 1.
DR DR SMART; SM00388; HsKA; 1.
KW Receptor.
SQ SEQUENCE 645 AA; 72192 MW; D37ABE3A0939D6ED CRC64;
Query Match 11.9%; Score 82.5; DB 2; Length 645;
Best Local Similarity 23.5%; Pred. NO. 17;
Matches 31; Conservative 26; Mismatches 46; Indels 29; Gaps 5;
QY 25 LHRLSNF---PRLPGILAPETV-----LPPCYKVKRKEKVKRSQ 63
DB 99 MTHLLAGTYGPHWPWMTAVTFKMLTGIVSFLTALSFLVTLPLLLKA--KVREFMLSK 156
QY 64 KATEFIDYSEIOSHHAILPLOTHTMKGSSMKC--SSLSEAILFTLTQLTOTLGLGCC 122
DB 157 KYRE-----LDREVGLMKQETSLHVRMLTYKIRSLDRHTLYTLVELSKTLGLKNC 211
QY 123 LLYLSKTIHPQI 134
DB 212 AVWIPNEIKTEM 223
RESULT 4
Q7SZX8 PRELIMINARY; PRT; 451 AA.
ID Q7SZX8
AC Q7SZX8;
DT 01-OCT-2003 (TEMBLrel. 25, Created)
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE SI:zk13421.9 (Novel protein similar to vertebrate microphthalmia-
DE associated transcription factor (MTF) and zebrafish transcription
DE factor binding to IGHM enhancer 3a (Tfe3a)).
DE Name=opnlh2; Synonyms=SI:zk13421.9;
DE OS Brachydanio rerio (Zebrafish) (Danio rerio).
DE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
DE Cyprinidae; Danio.
DE NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pandian R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC EMBL; AL844847; CAE30419.1; -.
DR HSSP; P22415; IAN4.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.

```